Adaptive estimation of spectral densities via wavelet thresholding and information projection

Jérémie Bigot, Rolando Biscay Lirio, Jean-Michel Loubes, Lilian Muniz Alvarez

Adaptive estimation of spectral densities via wavelet thresholding and information projection

Jérémie Bigot[1], Rolando J. Biscay Lirio[3], Jean-Michel Loubes[1] & Lilian Muñiz Alvarez[1,2]
IMT, Université Paul Sabatier, Toulouse, France[1]
Facultad de Matemática y Computación, Universidad de La Habana, Cuba[2]
CIMFAV-DEUV, Facultad de Ciencias, Universidad de Valparaiso, Chile[3]

June 6, 2011

Abstract

In this paper, we study the problem of adaptive estimation of the spectral density of a stationary Gaussian process. For this purpose, we consider a wavelet-based method which combines the ideas of wavelet approximation and estimation by information projection in order to warrants that the solution is a non-negative function. The spectral density of the process is estimated by projecting the wavelet thresholding expansion of the periodogram onto a family of exponential functions. This ensures that the spectral density estimator is a strictly positive function. The theoretical behavior of the estimator is established in terms of rate of convergence of the Kullback-Leibler discrepancy over Besov classes. We also show the excellent practical performance of the estimator in some numerical experiments.

Keywords: **Spectral density estimation, adaptive estimation, wavelet thresholding, sequences of exponential families, Besov spaces.**
AMS classifications: Primary 62G07; secondary 42C40, 41A29

1 Introduction

The estimation of spectral densities is a fundamental problem in inference for stationary stochastic processes. Many applications in several fields such as weather forecast and financial series are deeply related to this issue, see for instance Priestley [20]. It is known that the estimation of the covariance function of a stationary process is strongly related to the estimation of the corresponding spectral density. By Bochner's theorem the covariance function is non-negative definite if and only if the corresponding spectral density is a non-negative function. Hence in order to preserve the property of non-negative definiteness of a covariance function, the estimation of the corresponding spectral density must be a non-negative function. The purpose of this work is to provide a non-negative estimator of the spectral density.

Inference in the spectral domain uses the periodogram of the data, providing an inconsistent estimator which must be smoothed in order to achieve consistency. For highly regular spectral densities, linear smoothing techniques such as kernel smoothing are appropriate (see Brillinger [4]). However, these methods are not able to achieve the optimal mean-square rate of convergence for spectra whose smoothness is distributed inhomogeneously over the domain of interest. For this nonlinear methods are needed. One nonlinear method for adaptive spectral density estimation of a stationary Gaussian sequence was proposed by Comte [7]. It is based on model selection techniques. Others nonlinear smoothing procedures are the wavelet thresholding methods, first proposed by Donoho and Johnstone [12]. In this context, different thresholding rules have been proposed by Neumann [18] and Fryzlewicz, Nason and von Sachs [13] to name but a few.

Neumann's approach [18] consists in pre-estimating the variance of the periodogram via kernel smoothing, so that it can be supplied to the wavelet estimation procedure. Kernel pre-estimation may not be appropriate in cases where the underlying spectral density is of low regularity. One way to avoid this problem is proposed in Fryzlewicz, Nason and von Sachs [13], where the empirical wavelet coefficient thresholds are built as appropriate local weighted l_1 norms of the periodogram. These methods do not produce non-negative spectral density estimators, therefore the corresponding estimators of the covariance function is not non-negative definite.

To overcome the drawbacks of previous estimators, in this paper we propose a new wavelet-based method for the estimation of the spectral density of a Gaussian process. As a solution to ensure non-negativeness of the spectral density estimator, our method combines the ideas of wavelet thresholding and estimation by information projection. We estimate the spectral density by a projection of the nonlinear wavelet approximation of the periodogram onto a family of exponential functions. Therefore, the estimator is non-negative by construction. This technique was studied by Barron and Sheu [2] for the approximation of density functions by sequences of exponential families, by Loubes and Yan [16] for penalized maximum likelihood estimation with l_1 penalty, by Antoniadis and Bigot [1] for the study of Poisson inverse problems, and by Bigot and Van Bellegem [5] for log-density deconvolution.

The theoretical optimality of the estimators for the spectral density of a stationary process is generally studied using risk bounds in L_2-norm. This is the case in the papers of Neumann [18], Comte [7] and Fryzlewicz, Nason and von Sachs [13] mentioned before. In this work, the behavior of the proposed estimator is established in terms of the rate of convergence of the Kullback-Leibler discrepancy over Besov classes, which is maybe a more natural loss function for the estimation of a spectral density function than the L_2-norm. Moreover, the thresholding rules that we use to derive adaptive estimators differ from previous approaches based on wavelet decomposition and are quite simple to compute. Finally, we compare the performance of our estimator with other estimators on some simulations.

The paper is organized as follows. Section 2 presents the statistical framework under which we work. We define the model, the wavelet-based exponential family and the linear and nonlinear wavelet estimators by information projection. We also recall the definition of the Kullback-Leibler divergence and some results on Besov spaces. The rate of convergence of the proposed estimators are stated in Section 3. Some numerical experiments are described in Section 4. Technical lemmas and proofs of the main theorems are gathered in the Appendix.

Throughout this paper C denotes a constant that may vary from line to line. The notation $C(.)$ specifies the dependency of C on some quantities.

2 Statistical framework

2.1 The model

We aim at providing a nonparametric adaptive estimation of the spectral density which satisfies the property of being non-negative in order to guarantee that the covariance estimator is a non-negative definite function. We consider the sequence $(X_t)_{t\in\mathbb{N}}$ that satisfies the following assumptions:

Assumption 1 *The sequence $(X_1, ...X_n)$ is an n-sample drawn from a stationary sequence of Gaussian random variables.*

Let ρ be the covariance function of the process, i.e. $\rho(h) = cov(X_t, X_{t+h})$ with $h \in \mathbb{Z}$. The spectral density f is defined as:

$$f(\omega) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \rho(h) e^{i2\pi\omega h}, \ \omega \in [0,1].$$

We need the following standard assumption on ρ:

Assumption 2 *The covariance function ρ is non-negative definite, such that there exists two constants $0 < C_1, C_2 < +\infty$ such that $\sum_{h \in \mathbb{Z}} |\rho(h)| = C_1$ and $\sum_{h \in \mathbb{Z}} |h\rho^2(h)| = C_2$.*

Assumption 2 implies in particular that the spectral density f is bounded by the constant C_1. As a consequence, it is also square integrable. As in Comte [7], the data consist on a number of observations $X_1, ..., X_n$ at regularly spaced points. We want to obtain a positive estimator for the spectral density function f without parametric assumptions on the basis of these observations. For this, we combine the ideas of wavelet thresholding and estimation by information projection.

2.2 Estimation by information projection

To ensure nonnegativity of the estimator, we will look for approximations over an exponential family. For this, we construct a sieve of exponential functions defined in a wavelet basis.

Let $\phi(\omega)$ and $\psi(\omega)$, respectively, be the scaling and the wavelet functions generated by an orthonormal multiresolution decomposition of $L_2([0,1])$, see Mallat [17] for a detailed exposition on wavelet analysis. Throughout the paper, the functions ϕ and ψ are supposed to be compactly supported and such that $\|\phi\|_\infty < +\infty$, $\|\psi\|_\infty < +\infty$. Then, for any integer $j_0 \geq 0$, any function $g \in L_2([0,1])$ has the following representation:

$$g(\omega) = \sum_{k=0}^{2^{j_0}-1} \langle g, \phi_{j_0,k} \rangle \phi_{j_0,k}(\omega) + \sum_{j=j_0}^{+\infty} \sum_{k=0}^{2^j-1} \langle g, \psi_{j,k} \rangle \psi_{j,k}(\omega),$$

where $\phi_{j_0,k}(\omega) = 2^{\frac{j_0}{2}} \phi\left(2^{j_0}\omega - k\right)$ and $\psi_{j,k}(\omega) = 2^{\frac{j}{2}} \psi\left(2^j\omega - k\right)$. The main idea of this paper is to expand the spectral density f onto this wavelet basis and to find an estimator of this expansion that is then modified to impose the positivity property. The scaling and wavelet coefficients of the spectral density function f are denoted by $a_{j_0,k} = \langle f, \phi_{j_0,k} \rangle$ and $b_{j,k} = \langle f, \psi_{j,k} \rangle$.

To simplify the notations, we write $(\psi_{j,k})_{j=j_0-1}$ for the scaling functions $(\phi_{j,k})_{j=j_0}$. Let $j_1 \geq j_0$ and define the set

$$\Lambda_{j_1} = \left\{(j,k) : j_0 - 1 \leq j < j_1, 0 \leq k \leq 2^j - 1\right\}.$$

Note that $\#\Lambda_{j_1} = 2^{j_1}$, where $\#\Lambda_{j_1}$ denotes the cardinal of Λ_{j_1}. Let θ denotes a vector in $\mathbb{R}^{\#\Lambda_{j_1}}$, the wavelet-based exponential family $\mathcal{E}_{j_1}$ at scale j_1 is defined as the set of functions:

$$\mathcal{E}_{j_1} = \left\{ f_{j_1,\theta}(.) = \exp\left(\sum_{(j,k) \in \Lambda_{j_1}} \theta_{j,k} \psi_{j,k}(.) \right), \ \theta = (\theta_{j,k})_{(j,k) \in \Lambda_{j_1}} \in \mathbb{R}^{\#\Lambda_{j_1}} \right\}. \tag{2.1}$$

It is well known that Besov spaces for periodic functions in $L_2([0,1])$ can be characterized in terms of wavelet coefficients (see e.g. Mallat [17]). Assume that ψ has m vanishing moments, and let $0 < s < m$ denote the usual smoothness parameter. Then, for a Besov ball $B^s_{p,q}(A)$ of radius $A > 0$ with $1 \leq p, q \leq \infty$, one has that for $s^* = s + 1/2 - 1/p \geq 0$:

$$B^s_{p,q}(A) := \left\{ g \in L_2([0,1]) : \|g\|_{s,p,q} := \left(\sum_{k=0}^{2^{j_0}-1} |a_{j_0,k}|^p \right)^{\frac{1}{p}} + \left(\sum_{j=j_0}^{\infty} 2^{js^*q} \left(\sum_{k=0}^{2^j-1} |b_{j,k}|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}} \leq A \right\},$$

with the respective above sums replaced by maximum if $p = \infty$ or $q = \infty$ and where $a_{j_0,k} = \langle g, \phi_{j_0,k} \rangle$ and $b_{j,k} = \langle g, \psi_{j,k} \rangle$.

The condition that $s + 1/2 - 1/p \geq 0$ is imposed to ensure that $B^s_{p,q}(A)$ is a subspace of $L_2([0,1])$, and we shall restrict ourselves to this case in this paper (although not always stated, it is clear that all our results hold for $s < m$).

Let $M > 0$ and denote by $F^s_{p,q}(M)$ the set of functions such that

$$F^s_{p,q}(M) = \{ f = \exp(g) : \ \|g\|_{s,p,q} \leq M \},$$

where $\|g\|_{s,p,q}$ denotes the norm in the Besov space $B^s_{p,q}$. Note that assuming that $f \in F^s_{p,q}(M)$ implies that f is strictly positive. The following results hold.

Lemma 2.1 *Suppose that $f \in F^s_{p,q}(M)$ with $s > \frac{1}{p}$ and $1 \leq p \leq 2$. Then, there exists a constant M_1 such that for all $f \in F^s_{p,q}(M)$, $0 < M_1^{-1} \leq f \leq M_1 < +\infty$.*

Let V_j denote the usual multiresolution space at scale j spanned by the scaling functions $(\phi_{j,k})_{0 \leq k \leq 2^j - 1}$, and define $A_j < +\infty$ as the constant such that $\|v\|_\infty \leq A_j \|v\|_{L_2}$ for all $v \in V_j$. For $f \in F^s_{p,q}(M)$, let $g = \log(f)$. Then for $j \geq j_0 - 1$, define $D_j = \|g - g_j\|_{L_2}$ and $\gamma_j = \|g - g_j\|_\infty$, where $g_j = \sum_{k=0}^{2^j-1} \theta_{j,k}\psi_{j,k}$, with $\theta_{j,k} = \langle g, \psi_{j,k} \rangle$.

The proof of the following lemma immediately follows from the arguments in the proof of Lemma A.5 in Antoniadis and Bigot [1].

Lemma 2.2 *Let $j \in \mathbb{N}$. Then $A_j \leq C 2^{j/2}$. Suppose that $f \in F^s_{p,q}(M)$ with $1 \leq p \leq 2$ and $s > \frac{1}{p}$. Then, uniformly over $F^s_{p,q}(M)$, $D_j \leq C 2^{-j(s+1/2-1/p)}$ and $\gamma_j \leq C 2^{-j(s-1/p)}$ where C denotes constants depending only on M, s, p and q.*

To assess the quality of the estimators, we will measure the discrepancy between an estimator $\widehat{f}$ and the true function f in the sense of relative entropy (Kullback-Leibler divergence) defined by:

$$\Delta\left(f; \widehat{f}\right) = \int_0^1 \left(f \log\left(\frac{f}{\widehat{f}}\right) - f + \widehat{f} \right) d\mu,$$

where μ denotes the Lebesgue measure on $[0,1]$. It can be shown that $\Delta\left(f; \widehat{f}\right)$ is non-negative and equals zero if and only if $\widehat{f} = f$.

We will enforce our estimator of the spectral density to belong to the family $\mathcal{E}_{j_1}$ of exponential functions, which are positive by definition. For this we will consider a notion of projection using information projection.

The estimation of density function based on information projection has been introduced by Barron and Sheu [2]. To apply this method in our context, we recall for completeness a set of results that are useful to prove the existence of our estimators. The proofs of the following lemmas immediately follow from results in Barron and Sheu [2] and Antoniadis and Bigot [1].

Lemma 2.3 *Let $\beta \in \mathbb{R}^{\#\Lambda_{j_1}}$. Assume that there exists some $\theta(\beta) \in \mathbb{R}^{\#\Lambda_{j_1}}$ such that, for all $(j,k) \in \Lambda_{j_1}$, $\theta(\beta)$ is a solution of*

$$\left\langle f_{j,\theta(\beta)}, \psi_{j,k} \right\rangle = \beta_{j,k}.$$

Then for any function f such that $\langle f, \psi_{j,k} \rangle = \beta_{j,k}$ for all $(j,k) \in \Lambda_{j_1}$, and for all $\theta \in \mathbb{R}^{\#\Lambda_{j_1}}$, the following Pythagorian-like identity holds:

$$\Delta(f; f_{j,\theta}) = \Delta\left(f; f_{j,\theta(\beta)}\right) + \Delta\left(f_{j,\theta(\beta)}; f_{j,\theta}\right). \tag{2.2}$$

The next lemma is a key result which gives sufficient conditions for the existence of the vector $\theta(\beta)$ as defined in Lemma 2.3. This lemma also relates distances between the functions in the exponential family to distances between the corresponding wavelet coefficients. Its proof relies upon a series of lemmas on bounds within exponential families for the Kullback-Leibler divergence and can be found in Barron and Sheu [2] and Antoniadis and Bigot [1].

Lemma 2.4 *Let $\theta_0 \in \mathbb{R}^{\#\Lambda_{j_1}}$, $\beta_0 = \left(\beta_{0,(j,k)}\right)_{(j,k)\in\Lambda_{j_1}} \in \mathbb{R}^{\#\Lambda_{j_1}}$ such that $\beta_{0,(j,k)} = \langle f_{j,\theta_0}, \psi_{j,k} \rangle$ for all $(j,k) \in \Lambda_{j_1}$, and $\widetilde{\beta} \in \mathbb{R}^{\#\Lambda_{j_1}}$ a given vector. Let $b = \exp\left(\|\log(f_{j,\theta_0})\|_\infty\right)$ and $e = \exp(1)$. If $\left\|\widetilde{\beta} - \beta_0\right\|_2 \leq \frac{1}{2ebA_{j_1}}$ then the solution $\theta\left(\widetilde{\beta}\right)$ of*

$$\langle f_{j_1,\theta}, \psi_{j,k} \rangle = \widetilde{\beta}_{j,k} \text{ for all } (j,k) \in \Lambda_{j_1}$$

exists and satisfies

$$\begin{aligned}
\left\|\theta\left(\widetilde{\beta}\right) - \theta_0\right\|_2 &\leq 2eb\left\|\widetilde{\beta} - \beta_0\right\|_2 \\
\left\|\log\left(\frac{f_{j_1,\theta(\beta_0)}}{f_{j_1,\theta(\widetilde{\beta})}}\right)\right\|_\infty &\leq 2ebA_{j_1}\left\|\widetilde{\beta} - \beta_0\right\|_2 \\
\Delta\left(f_{j_1,\theta(\beta_0)}; f_{j_1,\theta(\widetilde{\beta})}\right) &\leq 2eb\left\|\widetilde{\beta} - \beta_0\right\|_2^2,
\end{aligned}$$

where $\|\beta\|_2$ denotes the standard Euclidean norm for $\beta \in \mathbb{R}^{\#\Lambda_{j_1}}$.

Following Csiszár [8], it is possible to define the projection of a function f onto $\mathcal{E}_{j_1}$. If this projection exists, it is defined as the function $f_{j_1,\theta^*_{j_1}}$ in the exponential family $\mathcal{E}_{j_1}$ that is the closest to the true function f in the Kullback-Leibler sense, and is characterized as the unique function in the family $\mathcal{E}_{j_1}$ for which

$$\left\langle f_{j_1,\theta^*_{j_1}}, \psi_{j,k} \right\rangle = \langle f, \psi_{j,k} \rangle := \beta_{j,k} \text{ for all } (j,k) \in \Lambda_{j_1}.$$

Note that the notation $\beta_{j,k}$ is used to denote both the the scaling coefficients $a_{j_0,k}$ and the wavelet coefficients $b_{j,k}$.

Let

$$I_n(\omega) = \frac{1}{2\pi n} \sum_{t=1}^{n} \sum_{t'=1}^{n} \left(X_t - \overline{X}\right) \left(X_{t'} - \overline{X}\right)^* e^{i2\pi\omega(t-t')},$$

be the classical periodogram, where $\left(X_t - \overline{X}\right)^*$ denotes the conjugate transpose of $\left(X_t - \overline{X}\right)$ and $\overline{X} = \frac{1}{n} \sum\limits_{t=1}^{n} X_t$. The expansion of $I_n(\omega)$ onto the wavelet basis allows to obtain estimators of $a_{j_0,k}$ and $b_{j,k}$ given by

$$\widehat{a}_{j_0,k} = \int_0^1 I_n(\omega)\, \phi_{j_0,k}(\omega)\, d\omega \quad \text{and} \quad \widehat{b}_{j,k} = \int_0^1 I_n(\omega)\, \psi_{j,k}(\omega)\, d\omega. \tag{2.3}$$

It seems therefore natural to estimate the function f by searching for some $\widehat{\theta}_n \in \mathbb{R}^{\#\Lambda_{j_1}}$ such that

$$\left\langle f_{j_1,\widehat{\theta}_n}, \psi_{j,k} \right\rangle = \int_0^1 I_n(\omega)\, \psi_{j,k}(\omega)\, d\omega := \widehat{\beta}_{j,k} \text{ for all } (j,k) \in \Lambda_{j_1}, \tag{2.4}$$

where $\widehat{\beta}_{j,k}$ denotes both the estimation of the scaling coefficients $\widehat{a}_{j_0,k}$ and the wavelet coefficients $\widehat{b}_{j,k}$. The function $f_{j_1,\widehat{\theta}_n}$ is the spectral density positive linear estimator.

Similarly, the positive nonlinear estimator with hard thresholding is defined as the function $f^{HT}_{j_1,\widehat{\theta}_n,\xi}$ (with $\widehat{\theta}_n \in \mathbb{R}^{\#\Lambda_{j_1}}$) such that

$$\left\langle f^{HT}_{j_1,\widehat{\theta}_n,\xi}, \psi_{j,k} \right\rangle = \delta_\xi\left(\widehat{\beta}_{j,k}\right) \text{ for all } (j,k) \in \Lambda_{j_1}, \tag{2.5}$$

where δ_ξ denotes the hard thresholding rule defined by

$$\delta_\xi(x) = xI(|x| \geq \xi) \text{ for } x \in \mathbb{R},$$

where $\xi > 0$ is an appropriate threshold whose choice is discussed later on.

The existence of these estimators is questionable. Thus, in the next sections, some sufficient conditions are given for the existence of $f_{j_1,\widehat{\theta}_n}$ and $f^{HT}_{j_1,\widehat{\theta}_n,\xi}$ with probability tending to one as $n \to +\infty$. Even if an explicit expression for $\widehat{\theta}_n$ is not available, we use a numerical approximation of $\widehat{\theta}_n$, obtained via a gradient-descent algorithm with an adaptive step.

In this section we establish the rate of convergence of our estimators in terms of the Kullback-Leibler discrepancy over Besov classes.

We make the following assumption on the wavelet basis that guarantees that Assumption 2 holds uniformly over $F^s_{p,q}(M)$.

Assumption 3 *Let $M > 0$, $1 \leq p \leq 2$ and $s > 1/p$. For $f \in F^s_{p,q}(M)$ and $h \in \mathbb{Z}$, let $\rho(h) = \int_0^1 f(\omega)e^{-i2\pi\omega h}d\omega$, $C_1(f) := \sum\limits_{h\in\mathbb{Z}} |\rho(h)|$ and $C_2(f) := \sum\limits_{h\in\mathbb{Z}} \left|h\rho^2(h)\right|$. Then, the wavelet basis is such that there exists a constant M_* such that for all $f \in F^s_{p,q}(M)$,*

$$C_1(f) \leq M_* \text{ and } C_2(f) \leq M_*.$$

2.3 Linear estimation

The following theorem is the general result on the linear information projection estimator of the spectral density function. Note that the choice of the coarse level resolution level j_0 is of minor importance, and without loss of generality we take $j_0 = 0$ for the linear estimator $f_{j_1,\widehat{\theta}_n}$.

Theorem 2.5 *Assume that $f \in F^s_{2,2}(M)$ with $s > \frac{1}{2}$ and suppose that Assumptions 1, 2 and 3 are satisfied. Define $j_1 = j_1(n)$ as the largest integer such that $2^{j_1} \leq n^{\frac{1}{2s+1}}$. Then, with probability tending to one as $n \to +\infty$, the information projection estimator (2.4) exists and satisfies:*

$$\Delta\left(f; f_{j_1(n),\widehat{\theta}_n}\right) = \mathcal{O}_p\left(n^{-\frac{2s}{2s+1}}\right).$$

Moreover, the convergence is uniform over the class $F^s_{2,2}(M)$ in the sense that

$$\lim_{K\to+\infty}\lim_{n\to+\infty}\sup_{f\in F^s_{2,2}(M)} \mathbb{P}\left(n^{\frac{2s}{2s+1}}\Delta\left(f; f_{j_1(n),\widehat{\theta}_n}\right) > K\right) = 0.$$

This theorem provides the existence with probability tending to one of a linear estimator for the spectral density f given by $f_{j_1(n),\widehat{\theta}_{j_1(n)}}$. This estimator is strictly positive by construction. Therefore the corresponding estimator of the covariance function $\widehat{\rho}^L$ (which is obtained as the inverse Fourier transform of $f_{j_1(n),\widehat{\theta}_n}$) is a positive definite function by Bochner's theorem. Hence $\widehat{\rho}^L$ is a covariance function.

In the related problem of density estimation from an i.d.d. sample, Koo [15] has shown that, for the Kullback-Leibler divergence, $n^{-\frac{2s}{2s+1}}$ is the fastest rate of convergence for the problem of estimating a density f such that $\log(f)$ belongs to the space $B^s_{2,2}(M)$. For spectral densities belonging to a general Besov ball $B^s_{p,q}(M)$, Newman [18] has also shown that $n^{-\frac{2s}{2s+1}}$ is an optimal rate of convergence for the L_2 risk. For the Kullback-Leibler divergence, we conjecture that $n^{-\frac{2s}{2s+1}}$ is the minimax rate of convergence for spectral densities belonging to $F^s_{2,2}(M)$.

However, the result obtained in the above theorem is nonadaptive because the selection of $j_1(n)$ depends on the unknown smoothness s of f. Moreover, the result is only suited for smooth functions (as $F^s_{2,2}(M)$ corresponds to a Sobolev space of order s) and does not attain an optimal rate of convergence when for example $g = \log(f)$ has singularities. We therefore propose in the next section an adaptive estimator derived by applying an appropriate nonlinear thresholding procedure.

2.4 Adaptive estimation

2.4.1 The bound on f is known

In adaptive estimation, we need to define an appropriate thresholding rule for the wavelet coefficients of the periodogram. This threshold is level-dependent and in this paper will take the form

$$\xi = \xi_{j,n} = 2\left[2\|f\|_\infty\left(\sqrt{\frac{\delta\log n}{n}} + 2^{\frac{j}{2}}\|\psi\|_\infty\frac{\delta\log n}{n}\right) + \frac{C_*}{\sqrt{n}}\right], \tag{2.6}$$

where $\delta \geq 0$ is a tuning parameter whose choice will be discussed later on and $C_* = \sqrt{\frac{C_2+39C_1^2}{4\pi^2}}$. The following theorem states that the relative entropy between the true f and its nonlinear estimator achieves in probability the conjectured optimal rate of convergence up to a logarithmic factor over a wide range of Besov balls.

Theorem 2.6 *Assume that $f \in F^s_{p,q}(M)$ with $s > \frac{1}{2} + \frac{1}{p}$ and $1 \le p \le 2$. Suppose also that Assumptions 1, 2, 3 hold. For any $n > 1$, define $j_0 = j_0(n)$ to be the integer such that $2^{j_0} \ge \log n \ge 2^{j_0 - 1}$, and $j_1 = j_1(n)$ to be the integer such that $2^{j_1} \ge \frac{n}{\log n} \ge 2^{j_1-1}$. For $\delta \ge 6$, take the threshold $\xi_{j,n}$ as in (2.6). Then, the thresholding estimator (2.5) exists with probability tending to one when $n \to +\infty$ and satisfies:*

$$\Delta\left(f; f^{HT}_{j_0(n),j_1(n),\widehat{\theta}_n,\xi_{j,n}}\right) = \mathcal{O}_p\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right).$$

Note that the choices of j_0, j_1 and $\xi_{j,n}$ are independent of the parameter s; hence the estimator $f^{HT}_{j_0(n),j_1(n),\widehat{\theta}_n,\xi_{j,n}}$ is an adaptive estimator which attains in probability what we claim is the optimal rate of convergence, up to a logarithmic factor. In particular, $f^{HT}_{j_0(n),j_1(n),\widehat{\theta}_n,\xi_{j,n}}$ is adaptive on $F^s_{2,2}(M)$. This theorem provides the existence with probability tending to one of a nonlinear estimator for the spectral density. This estimator is strictly positive by construction. Therefore the corresponding estimator of the covariance function $\widehat{\rho}^{NL}$ (which is obtained as the inverse Fourier transform of $f^{HT}_{j_0(n),j_1(n),\widehat{\theta}_n,\xi_{j,n}}$) is a positive definite function by Bochner theorem. Hence $\widehat{\rho}^{NL}$ is a covariance function.

2.4.2 Estimating the bound on f

Although the results of Theorem 2.6 are certainly of some theoretical interest, they are not helpful for practical applications. The (deterministic) threshold $\xi_{j,n}$ depends on the unknown quantities $\|f\|_\infty$ and $C_* := C(C_1, C_2)$, where C_1 and C_2 are unknown constants. To make the method applicable, it is necessary to find some completely data-driven rule for the threshold, which works well over a range as wide as possible of smoothness classes. In this subsection, we give an extension that leads to consider a random threshold which no longer depends on the bound on f neither on C_*. For this let us consider the dyadic partitions of $[0,1]$ given by $\mathcal{I}_n = \left\{\left(j/2^{J_n}, (j+1)/2^{J_n}\right),\ j = 0, ..., 2^{J_n} - 1\right\}$. Given some positive integer r, we define $\mathcal{P}_n$ as the space of piecewise polynomials of degree r on the dyadic partition $\mathcal{I}_n$ of step 2^{-J_n}. The dimension of $\mathcal{P}_n$ depends on n and is denoted by N_n. Note that $N_n = (r+1)\,2^{J_n}$. This family is regular in the sense that the partition $\mathcal{I}_n$ has equispaced knots.

An estimator of $\|f\|_\infty$ is constructed as proposed by Birgé and Massart [6] in the following way. We take the infinite norm of $\widehat{f}_n$, where $\widehat{f}_n$ denotes the (empirical) orthogonal projection of the periodogram I_n on $\mathcal{P}_n$. We denote by f_n the L_2-orthogonal projection of f on the same space. Then the following theorem holds.

Theorem 2.7 *Assume that $f \in F^s_{p,q}(M)$ with $s > \frac{1}{2} + \frac{1}{p}$ and $1 \le p \le 2$. Suppose also that Assumptions 1, 2 and 3 hold. For any $n > 1$, let $j_0 = j_0(n)$ be the integer such that $2^{j_0} \ge \log n \ge 2^{j_0-1}$, and let $j_1 = j_1(n)$ be the integer such that $2^{j_1} \ge \frac{n}{\log n} \ge 2^{j_1-1}$. Take the constants $\delta = 6$ and $b \in \left[\frac{3}{4}, 1\right)$, and define the threshold*

$$\widehat{\xi}_{j,n} = 2\left[2\left\|\widehat{f}_n\right\|_\infty\left(\sqrt{\frac{\delta}{(1-b)^2}\frac{\log n}{n}} + 2^{\frac{j}{2}}\|\psi\|_\infty\frac{\delta}{(1-b)^2}\frac{\log n}{n}\right) + \sqrt{\frac{\log n}{n}}\right]. \quad (2.7)$$

Then, if $\|f - f_n\|_\infty \le \frac{1}{4}\|f\|_\infty$ and $N_n \le \frac{\kappa}{(r+1)^2}\frac{n}{\log n}$, where κ is a numerical constant and r is the degree of the polynomials, the thresholding estimator (2.5) exists with probability tending to one as

$n \to +\infty$ and satisfies

$$\Delta \left(f; f^{HT}_{j_0(n),j_1(n),\widehat{\theta}_n,\widehat{\xi}_{j,n}} \right) = \mathcal{O}_p \left(\left(\frac{n}{\log n} \right)^{-\frac{2s}{2s+1}} \right).$$

Note that, we finally obtain a fully tractable estimator of f which reaches the optimal rate of convergence without prior knowledge of the regularity of the spectral density, but also which gets rise to a real covariance estimator.

Remark 2.8 *We point out that, in Comte [7] the condition $\|f - f_n\|_\infty \leq \frac{1}{4} \|f\|_\infty$ is assumed. Under some regularity conditions on f, results from approximation theory entails that this condition is met. Indeed for $f \in B^s_{p,\infty}$, with $s > \frac{1}{p}$, we know from DeVore and Lorentz [11] that*

$$\|f - f_n\|_\infty \leq C(s) |f|_{s,p} N_n^{-\left(s - \frac{1}{p}\right)},$$

with $|f|_{s,p} = \sup_{y>0} y^{-s} w_d (f,y)_p < +\infty$, where $w_d(f,y)_p$ is the modulus of smoothness and $d = [s] + 1$. Therefore $\|f - f_n\|_\infty \leq \frac{1}{4} \|f\|_\infty$ if $N_n \geq \left(4C(s) \frac{|f|_{s,p}}{\|f\|_\infty} \right)^{\frac{1}{s - \frac{1}{p}}} := C(f,s,p)$, where $C(f,s,p)$ is a constant depending on f, s and p.

3 Numerical experiments

In this section we present some numerical experiments which support the claims made in the theoretical part of this paper. The programs for our simulations were implemented using the MATLAB programming environment. We simulate a time series which is a superposition of an ARMA(2,2) process and a Gaussian white noise:

$$X_t = Y_t + c_o Z_t, \tag{3.1}$$

where $Y_t + a_1 Y_{t-1} + a_2 Y_{t-2} = b_0 \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2}$, and $\{\varepsilon_t\}$, $\{Z_t\}$ are independent Gaussian white noise processes with unit variance. The constants were chosen as $a_1 = 0.2$, $a_2 = 0.9$, $b_0 = 1$, $b_1 = 0$, $b_2 = 1$ and $c_0 = 0.5$. We generated a sample of size $n = 1024$ according to (3.1). The spectral density f of (X_t) is shown in Figure 1. It has two moderately sharp peaks and is smooth in the rest of the domain.

Starting from the periodogram we considered the Symmlet 8 basis, i.e. the least asymmetric, compactly supported wavelets which are described in Daubechies [9]. We choose j_0 and j_1 as in the hypothesis of Theorem 2.7 and left the coefficients assigned to the father wavelets unthresholded. Hard thresholding is performed using the threshold $\widehat{\xi}_{j,n}$ as in (2.7) for the levels $j = j_0, ..., j_1$, and the empirical coefficients from the higher resolution scales $j > j_1$ are set to zero. This gives the estimate

$$f^{HT}_{j_0,j_1,\xi_{j,n}} = \sum_{k=0}^{2^{j_0}-1} \widehat{a}_{j_0,k} \phi_{j_0,k} + \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \widehat{b}_{j,k} I \left(\left| \widehat{b}_{j,k} \right| > \xi_{j,n} \right) \psi_{j,k}, \tag{3.2}$$

which is obtained by simply thresholding the wavelet coefficients (2.3) of the periodogram. Note that such an estimator is not guaranteed to be strictly positive in the interval $[0,1]$. However, we use it to built our strictly positive estimator $f^{HT}_{j_0,j_1,\widehat{\theta}_n,\widehat{\xi}_{j,n}}$ (see (2.5) to recall its definition). We want to find $\widehat{\theta}_n$ such that

$$\left\langle f^{HT}_{j_0,j_1,\widehat{\theta}_n,\widehat{\xi}_{j,n}}, \psi_{j,k} \right\rangle = \delta_{\widehat{\xi}_{j,n}}\left(\widehat{\beta}_{j,k}\right) \text{ for all } (j,k) \in \Lambda_{j_1}$$

For this, we take

$$\widehat{\theta}_n = \underset{\theta\in\mathbb{R}^{\#\Lambda_{j_1}}}{\arg\min} \sum_{(j,k)\in\Lambda_{j_1}} \left(\langle f_{j_0,j_1,\theta}, \psi_{j,k} \rangle - \delta_{\widehat{\xi}_{j,n}}\left(\widehat{\beta}_{j,k}\right)\right)^2,$$

where $f_{j_0,j_1,\theta}(.) = \exp\left(\sum_{(j,k)\in\Lambda_{j_1}} \theta_{j,k}\psi_{j,k}(.)\right) \in \mathcal{E}_{j_1}$ and $\mathcal{E}_{j_1}$ is the family (2.1). To solve this optimization problem we used a gradient descent method with an adaptive step, taking as initial value

$$\theta_0 = \left\langle \log\left(\left(f^{HT}_{j_0,j_1,\widehat{\xi}_{j,n}}\right)_+\right), \psi_{j,k} \right\rangle,$$

where $\left(f^{HT}_{j_0,j_1,\widehat{\xi}_{j,n}}(\omega)\right)_+ := \max\left(f^{HT}_{j_0,j_1,\widehat{\xi}_{j,n}}(\omega), \eta\right)$ for all $\omega \in [0,1]$ and $\eta > 0$ is a small constant.

In Figure 1 we display the unconstrained estimator $f^{HT}_{j_0,j_1,\xi_{j,n}}$ as in (3.2), obtained by thresholding of the wavelet coefficients of the periodogram, together with the estimator $f^{HT}_{j_0,j_1,\widehat{\theta}_n,\widehat{\xi}_{j,n}}$, which is strictly positive by construction. Note that these wavelet estimators capture well the peaks and look fairly good on the smooth part too.

We compared our method with the spectral density estimator proposed by Comte [7], which is based on a model selection procedure. As an example, in Comte [7], the author study the behavior of such estimators using a collection of nested models (S_m), with $m = 1, ..., 100$, where S_m is the space of piecewise constant functions, generated by a histogram basis on $[0,1]$ of dimension m with equispaced knots (see Comte [7] for further details). In Figure 2 we show the result of this comparison. Note that our method better captures the peaks of the true spectral density.

Acknowledgments

This work was supported in part by Egide, under the Program of Eiffel excellency Phd grants, as well as by the BDI CNRS grant.

4 Appendix

Throughout all the proofs, C denotes a generic constant whose value may change from line to line.

4.1 Technical results for the empirical estimators of the wavelet coefficients

Lemma 4.1 *Let $n \geq 1$, $\beta_{j,k} := \langle f, \psi_{j,k} \rangle$ and $\widehat{\beta}_{j,k} := \langle I_n, \psi_{j,k} \rangle$ for $j \geq j_0 - 1$ and $0 \leq k \leq 2^j - 1$. Suppose that Assumptions 1, 2 and 3 hold. Then, $Bias^2\left(\widehat{\beta}_{j,k}\right) := \left(\mathbb{E}\left(\widehat{\beta}_{j,k}\right) - \beta_{j,k}\right)^2 \leq \frac{C_*^2}{n}$ with*

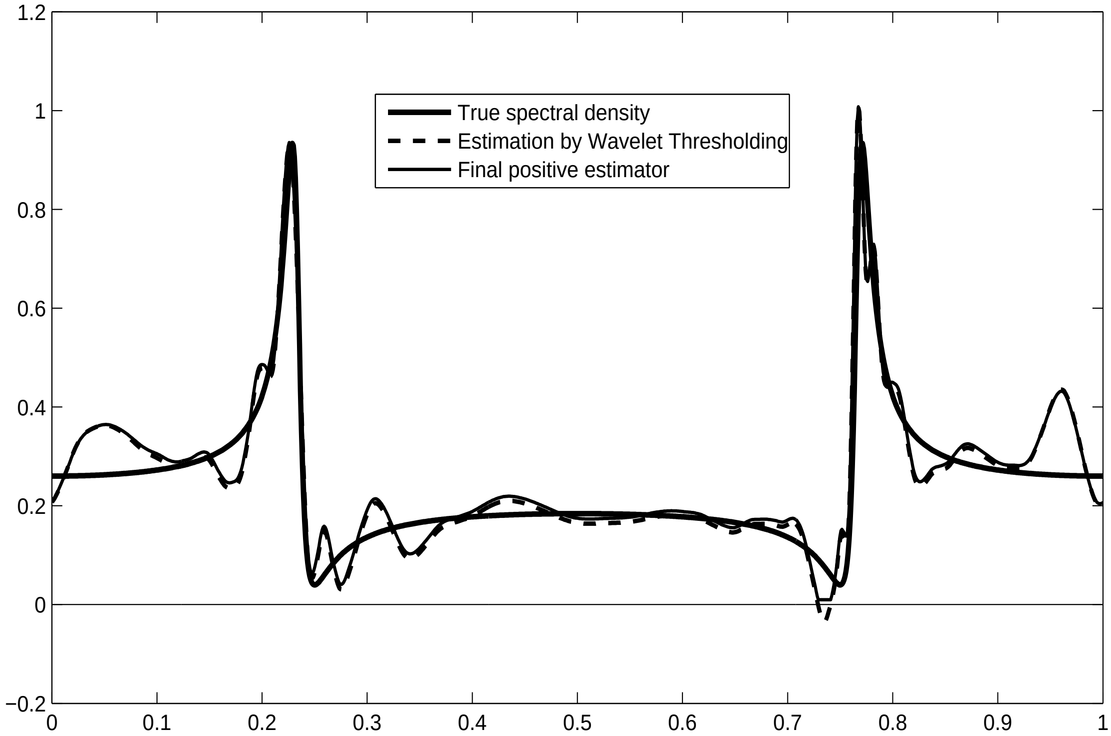

Figure 1: True spectral density f, wavelet thresholding estimator $f^{HT}_{j_0,j_1,\widehat{\xi}_{j,n}}$ and final positive estimator $f^{HT}_{j_0,j_1,\widehat{\theta}_n,\widehat{\xi}_{j,n}}$.

$C_* = \sqrt{\frac{C_2+39C_1^2}{4\pi^2}}$, *and* $Var\left(\widehat{\beta}_{j,k}\right) := \mathbb{E}\left(\widehat{\beta}_{j,k} - \mathbb{E}\left(\widehat{\beta}_{j,k}\right)\right)^2 \leq \frac{C}{n}$ *for some constant* $C > 0$. *Moreover, there exists a constant* $M_2 > 0$ *such that for all* $f \in F^s_{p,q}(M)$ *with* $s > \frac{1}{p}$ *and* $1 \leq p \leq 2$,

$$\mathbb{E}\left(\widehat{\beta}_{j,k} - \beta_{j,k}\right)^2 = Bias^2\left(\widehat{\beta}_{j,k}\right) + Var\left(\widehat{\beta}_{j,k}\right) \leq \frac{M_2}{n}.$$

Proof. Note that $Bias^2\left(\widehat{\beta}_{j,k}\right) \leq \|f - \mathbb{E}\left(I_n\right)\|^2_{L_2}$. Using Proposition 1 in Comte [7], Assumptions 1 and 2 imply that $\|f - \mathbb{E}\left(I_n\right)\|^2_{L_2} \leq \frac{C_2+39C_1^2}{4\pi^2 n}$, which gives the result for the bias term. To bound the variance term, remark that

$$Var\left(\widehat{\beta}_{j,k}\right) = \mathbb{E}\left\langle I_n - \mathbb{E}\left(I_n\right), \psi_{j,k}\right\rangle^2 \leq \mathbb{E}\|I_n - \mathbb{E}\left(I_n\right)\|^2_{L_2}\|\psi_{j,k}\|^2_{L_2} = \int_0^1 \mathbb{E}|I_n(\omega) - \mathbb{E}\left(I_n(\omega)\right)|^2 d\omega.$$

Then, under Assumptions 1 and 2, it follows that there exists an absolute constant $C > 0$ such that for all $\omega \in [0,1]$, $\mathbb{E}|I_n(\omega) - \mathbb{E}\left(I_n(\omega)\right)|^2 \leq \frac{C}{n}$. To complete the proof it remains to remark that Assumption 3 implies that these bounds for the bias and the variance hold uniformly over $F^s_{p,q}(M)$. ∎

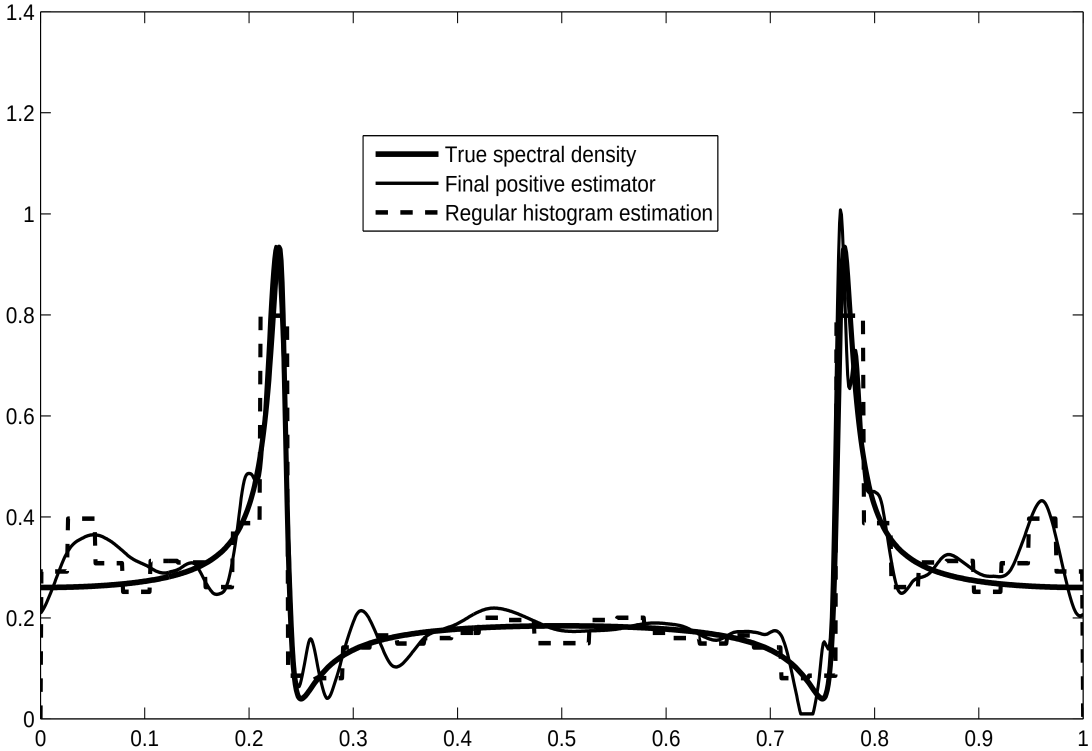

Figure 2: True spectral density f, final positive estimator $f^{HT}_{j_0,j_1,\widehat{\theta}_n,\widehat{\xi}_{j,n}}$ and estimator via model selection using regular histograms.

Lemma 4.2 *Let $n \geq 1$, $b_{j,k} := \langle f, \psi_{j,k} \rangle$ and $\widehat{b}_{j,k} := \langle I_n, \psi_{j,k} \rangle$ for $j \geq j_0$ and $0 \leq k \leq 2^j - 1$. Suppose that Assumptions 1 and 2 hold. Then for any $x > 0$,*

$$\mathbb{P}\left(|\widehat{b}_{j,k} - b_{j,k}| > 2\|f\|_\infty \left(\sqrt{\frac{x}{n}} + 2^{j/2}\|\psi\|_\infty \frac{x}{n} \right) + \frac{C_*}{\sqrt{n}} \right) \leq 2e^{-x},$$

where $C_* = \sqrt{\frac{C_2 + 39C_1^2}{4\pi^2}}$.

Proof. Note that

$$\widehat{b}_{j,k} = \frac{1}{2\pi n} \sum_{t=1}^{n} \sum_{t'=1}^{n} \left(X_t - \overline{X} \right) \left(X_{t'} - \overline{X} \right)^* \int_0^1 e^{i2\pi\omega(t-t')} \psi_{j,k}(\omega)\, d\omega = \frac{1}{2\pi n} X^T T_n(\psi_{j,k}) X^*,$$

where $X = \left(X_1 - \overline{X}, ..., X_n - \overline{X} \right)^T$, X^T denotes the transpose of X and $T_n(\psi_{j,k})$ is the Toeplitz matrix with entries $[T_n(\psi_{j,k})]_{t,t'} = \int_0^1 e^{i2\pi\omega(t-t')} \psi_{j,k}(\omega)\, d\omega$, $1 \leq t, t' \leq n$. We can assume without loss of generality that $E(X_t) = 0$, and then under under Assumptions 1 and 2, X is a centered

Gaussian vector in $\mathbb{R}^n$ with covariance matrix $\Sigma = T_n(f)$. Using the decomposition $X = \Sigma^{\frac{1}{2}}\varepsilon$, where $\varepsilon \sim N(0, I_n)$, it follows that $\widehat{b}_{j,k} = \frac{1}{2\pi n}\varepsilon^T A_{j,k}\varepsilon$, with $A_{j,k} = \Sigma^{\frac{1}{2}} T_n(\psi_{j,k})\Sigma^{\frac{1}{2}}$. Note also that $\mathbb{E}\left(\widehat{b}_{j,k}\right) = \frac{1}{2\pi n} tr\left(A_{j,k}\right)$, where $tr\left(A\right)$ denotes the trace of a matrix A.

Now let $s_1, \ldots, s_n$ be the eigenvalues of the Hermitian matrix $A_{j,k}$ with $|s_1| \geq |s_2| \geq \ldots \geq |s_n|$ and let $Z = 2\pi n\left(\widehat{b}_{j,k} - \mathbb{E}\left(\widehat{b}_{j,k}\right)\right) = \varepsilon^T A_{j,k}\varepsilon - tr\left(A_{j,k}\right)$. Then, for $0 < \lambda < (2|s_1|)^{-1}$ one has that

$$\begin{aligned}
\log\left(\mathbb{E}\left(e^{\lambda Z}\right)\right) &= \sum_{i=1}^{n} -\lambda s_i - \frac{1}{2}\log\left(1 - 2\lambda s_i\right) = \sum_{i=1}^{n}\sum_{\ell=2}^{+\infty}\frac{1}{2\ell}(2 s_i \lambda)^{\ell} \leq \sum_{i=1}^{n}\sum_{\ell=2}^{+\infty}\frac{1}{2\ell}(2|s_i|\lambda)^{\ell} \\
&\leq \sum_{i=1}^{n} -\lambda|s_i| - \frac{1}{2}\log\left(1 - 2\lambda|s_i|\right),
\end{aligned}$$

where we have used the fact that $-\log(1-x) = \sum_{\ell=1}^{+\infty}\frac{x^{\ell}}{\ell}$ for $x < 1$. Then using the inequality $-u - \frac{1}{2}\log\left(1 - 2u\right) \leq \frac{u^2}{1-2u}$ that holds for all $0 < u < \frac{1}{2}$, the above inequality implies that

$$\log\left(\mathbb{E}\left(e^{\lambda Z}\right)\right) \leq \sum_{i=1}^{n}\frac{\lambda^2 |s_i|^2}{1 - 2\lambda|s_i|} \leq \frac{\lambda^2\|s\|^2}{1 - 2\lambda|s_1|},$$

where $\|s\|^2 = \sum_{i=1}^{n}|s_i|^2$. Arguing as in Birgé and Massart [6], the above inequality implies that for any $x > 0$, $\mathbb{P}(|Z| > 2|s_1|x + 2\|s\|\sqrt{x}) \leq 2e^{-x}$, which implies

$$\mathbb{P}\left(|\widehat{b}_{j,k} - \mathbb{E}\left(\widehat{b}_{j,k}\right)| > 2|s_1|\frac{x}{n} + 2\frac{\|s\|}{n}\sqrt{x}\right) \leq 2e^{-x}. \tag{4.1}$$

Let $\tau\left(A\right)$ denotes the spectral radius of a matrix A. For the Toeplitz matrices $\Sigma = T_n(f)$ and $T_n(\psi_{j,k})$ one has that $\tau\left(\Sigma\right) \leq \|f\|_{\infty}$ and $\tau\left(T_n(\psi_{j,k})\right) \leq \|\psi_{j,k}\|_{\infty} = 2^{j/2}\|\psi\|_{\infty}$. These inequalities imply that

$$|s_1| = \tau\left(\Sigma^{\frac{1}{2}} T_n(\psi_{j,k})\Sigma^{\frac{1}{2}}\right) \leq \tau\left(\Sigma\right)\tau\left(T_n(\psi_{j,k})\right) \leq \|f\|_{\infty}2^{j/2}\|\psi\|_{\infty}. \tag{4.2}$$

Let λ_i, $i = 1.,,,.n$, be the eigenvalues of $T_n(\psi_{j,k})$. From Lemma 3.1 in Davies [10], we have that

$$\limsup_{n\to+\infty}\frac{1}{n} tr\left(T_n(\psi_{j,k})^2\right) = \limsup_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}\lambda_i^2 = \int_0^1 \psi_{j,k}^2\left(\omega\right)d\omega = 1,$$

which implies that

$$\|s\|^2 = \sum_{i=1}^{n}|s_i|^2 = tr\left(A_{j,k}^2\right) = tr\left(\left(\Sigma T_n(\psi_{j,k})\right)^2\right) \leq \tau\left(\Sigma\right)^2 tr\left(T_n(\psi_{j,k})^2\right) \leq \|f\|_{\infty}^2 n, \tag{4.3}$$

where we have used the inequality $tr\left((AB)^2\right) \leq \tau\left(A\right)^2 tr\left(B^2\right)$ that holds for any pair of Hermitian matrices A, B. Combining (4.1), (4.2) and (4.3), we finally obtain that for any $x > 0$

$$\mathbb{P}\left(|\widehat{b}_{j,k} - \mathbb{E}\left(\widehat{b}_{j,k}\right)| > 2\|f\|_{\infty}\left(\sqrt{\frac{x}{n}} + 2^{j/2}\|\psi\|_{\infty}\frac{x}{n}\right)\right) \leq 2e^{-x}. \tag{4.4}$$

Now, let $\xi_{j,n} = 2\|f\|_\infty \left(\sqrt{\frac{x}{n}} + 2^{j/2}\|\psi\|_\infty \frac{x}{n}\right) + \frac{C_*}{\sqrt{n}}$, and note that

$$\mathbb{P}\left(\left|\widehat{b}_{j,k} - b_{j,k}\right| > \xi_{j,n}\right) \leq \mathbb{P}\left(\left|\widehat{b}_{j,k} - \mathbb{E}\left(\widehat{b}_{j,k}\right)\right| > \xi_{j,n} - \left|\mathbb{E}\left(\widehat{b}_{j,k}\right) - b_{j,k}\right|\right),$$

By Lemma 4.1, one has that $\left|\mathbb{E}\left(\widehat{b}_{j,k}\right) - b_{j,k}\right| \leq \frac{C_*}{\sqrt{n}}$, and thus $\xi_{j,n} - \left|\mathbb{E}\left(\widehat{b}_{j,k}\right) - b_{j,k}\right| \geq \xi_{j,n} - \frac{C_*}{\sqrt{n}}$ which implies using (4.4) that

$$\mathbb{P}\left(\left|\widehat{b}_{j,k} - b_{j,k}\right| > \xi_{j,n}\right) \leq \mathbb{P}\left(\left|\widehat{b}_{j,k} - \mathbb{E}\left(\widehat{b}_{j,k}\right)\right| > \xi_{j,n} - \frac{C_*}{\sqrt{n}}\right) \leq 2e^{-x},$$

which completes the proof of Lemma 4.2. ∎

Lemma 4.3 *Assume that $f \in F^s_{p,q}(M)$ with $s > \frac{1}{2} + \frac{1}{p}$ and $1 \leq p \leq 2$. Suppose that Assumptions 1, 2 and 3 hold. For any $n > 1$, define $j_0 = j_0(n)$ to be the integer such that $2^{j_0} > \log n \geq 2^{j_0 - 1}$, and $j_1 = j_1(n)$ to be the integer such that $2^{j_1} \geq \frac{n}{\log n} \geq 2^{j_1 - 1}$. For $\delta \geq 6$, take the threshold $\xi_{j,n} = 2\left[2\|f\|_\infty\left(\sqrt{\frac{\delta \log n}{n}} + 2^{\frac{j}{2}}\|\psi\|_\infty \frac{\delta \log n}{n}\right) + \frac{C_*}{\sqrt{n}}\right]$ as in (2.6), where $C_* = \sqrt{\frac{C_2 + 39C_1^2}{4\pi^2}}$. Let $\beta_{j,k} := \langle f, \psi_{j,k}\rangle$ and $\widehat{\beta}_{\xi_{j,n},(j,k)} := \delta_{\xi_{j,n}}\left(\widehat{\beta}_{j,k}\right)$ with $(j,k) \in \Lambda_{j_1}$ as in (2.5). Take $\beta = (\beta_{j,k})_{(j,k)\in\Lambda_{j_1}}$ and $\widehat{\beta}_{\xi_{j,n}} = \left(\widehat{\beta}_{\xi_{j,n},(j,k)}\right)_{(j,k)\in\Lambda_{j_1}}$. Then there exists a constant $M_3 > 0$ such that for all sufficiently large n:*

$$\mathbb{E}\left\|\beta - \widehat{\beta}_{\xi_{j,n}}\right\|_2^2 := \mathbb{E}\left(\sum_{(j,k)\in\Lambda_{j_1}}\left|\beta_{j,k} - \delta_{\xi_{j,n}}\left(\widehat{\beta}_{j,k}\right)\right|^2\right) \leq M_3\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}$$

uniformly over $F^s_{p,q}(M)$.

Proof. Taking into account that

$$\begin{aligned}
\mathbb{E}\left\|\beta - \widehat{\beta}_{\xi_{j,n}}\right\|_2^2 &= \sum_{k=0}^{2^{j_0}-1}\mathbb{E}\left(a_{j_0,k} - \widehat{a}_{j_0,k}\right)^2 + \sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1}\mathbb{E}\left[\left(b_{j,k} - \widehat{b}_{j,k}\right)^2 I\left(\left|\widehat{b}_{j,k}\right| > \xi_{j,n}\right)\right] \\
&\quad + \sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} b_{j,k}^2 P\left(\left|\widehat{b}_{j,k}\right| \leq \xi_{j,n}\right) \\
&:= T_1 + T_2 + T_3, \qquad (4.5)
\end{aligned}$$

we are interested in bounding these three terms. The bound for T_1 follows from Lemma 4.1 and the fact that $j_0 = \log_2(\log n) \leq \frac{1}{2s+1}\log_2(n)$:

$$T_1 = \sum_{k=0}^{2^{j_0}-1}\mathbb{E}\left(a_{j_0,k} - \widehat{a}_{j_0,k}\right)^2 = O\left(\frac{2^{j_0}}{n}\right) \leq O\left(n^{-\frac{2s}{2s+1}}\right). \qquad (4.6)$$

To bound T_2 and T_3 we proceed as follows. Write

$$T_2 = \sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1}\mathbb{E}\left[\left(b_{j,k} - \widehat{b}_{j,k}\right)^2\left\{I\left(\left|\widehat{b}_{j,k}\right| > \xi_{j,n}, |b_{j,k}| > \frac{\xi_{j,n}}{2}\right) + I\left(\left|\widehat{b}_{j,k}\right| > \xi_{j,n}, |b_{j,k}| \leq \frac{\xi_{j,n}}{2}\right)\right\}\right]$$

and

$$T_3 = \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} b_{j,k}^2 \left[P\left(\left| \widehat{b}_{j,k} \right| \le \xi_{j,n}\,, |b_{j,k}| \le 2\xi_{j,n} \right) + P\left(\left| \widehat{b}_{j,k} \right| \le \xi_{j,n}, |b_{j,k}| > 2\xi_{j,n} \right) \right].$$

From Hardle, Kerkyacharian, Picard and Tsybakov [14] we get that

$$\begin{aligned} T_2 + T_3 \le & \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left\{ \left(b_{j,k} - \widehat{b}_{j,k} \right)^2 \right\} I\left(|b_{j,k}| > \frac{\xi_{j,n}}{2} \right) + \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} b_{j,k}^2 I\left(|b_{j,k}| \le 2\xi_{j,n} \right) \\ & + 5 \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left\{ \left(b_{j,k} - \widehat{b}_{j,k} \right)^2 I\left(\left| \widehat{b}_{j,k} - b_{j,k} \right| > \frac{\xi_{j,n}}{2} \right) \right\} \\ & := T' + T'' + T'''. \end{aligned}$$

Now we bound T'''. Using Cauchy-Schwarz inequality, we obtain

$$T''' \le 5 \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}^{\frac{1}{2}} \left[\left(b_{j,k} - \widehat{b}_{j,k} \right)^4 \right] P^{\frac{1}{2}} \left(\left| \widehat{b}_{j,k} - b_{j,k} \right| > \frac{\xi_{j,n}}{2} \right).$$

By the same inequality we get $\mathbb{E}\left[\left(\widehat{b}_{j,k} - b_{j,k} \right)^4 \right] \le \mathbb{E}\left[\|I_n - f\|_{L_2}^4 \|\psi_{j,k}\|_{L_2}^4 \right] = O\left(\mathbb{E}\|I_n - f\|_{L_2}^4 \right)$. It can be checked that $\mathbb{E}\|I_n - f\|_{L_2}^4 \le 8\mathbb{E}\left(\|I_n - \mathbb{E}(I_n)\|_{L_2}^4 + \|\mathbb{E}(I_n) - f\|_{L_2}^4 \right)$. According to Comte [7], $\mathbb{E}\|I_n - \mathbb{E}(I_n)\|_{L_2}^4 = O(n^2)$. From the proof of Lemma 4.1 we get that $\|\mathbb{E}(I_n) - f\|_{L_2}^4 = O\left(\frac{1}{n^2}\right)$. Therefore $\mathbb{E}\|I_n - f\|_{L_2}^4 \le O\left(n^2 + \frac{1}{n^2}\right) = O(n^2)$. Hence $\mathbb{E}\left[\left(\widehat{b}_{j,k} - b_{j,k} \right)^4 \right] = O\left(\mathbb{E}\|I_n - f\|_{L_2}^4 \right) = O(n^2)$. For the bound of $P\left(\left| \widehat{b}_{j,k} - b_{j,k} \right| > \frac{\xi_{j,n}}{2} \right)$ we use the result of Lemma 4.2 with $x = \delta \log n$, where $\delta > 0$ is a constant to be specified later. We obtain

$$\begin{aligned} P\left(\left| \widehat{b}_{j,k} - b_{j,k} \right| > \frac{\xi_{j,n}}{2} \right) &= P\left(\left| \widehat{b}_{j,k} - b_{j,k} \right| > 2\|f\|_\infty \left(\sqrt{\frac{\delta \log n}{n}} + 2^{\frac{j}{2}} \|\psi\|_\infty \frac{\delta \log n}{n} \right) + \frac{C_*}{\sqrt{n}} \right) \\ &\le 2e^{-\delta \log n} = 2n^{-\delta}. \end{aligned}$$

Therefore, for $\delta \ge 6$, we get

$$T''' \le 5 \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}^{\frac{1}{2}} \left[\left(b_{j,k} - \widehat{b}_{j,k} \right)^4 \right] P^{\frac{1}{2}} \left(\left| \widehat{b}_{j,k} - b_{j,k} \right| > \frac{\xi_{j,n}}{2} \right) \le O\left(\frac{n^{-1}}{\log n} \right) \le O\left(n^{-\frac{2s}{2s+1}} \right).$$

Now we follow results found in Pensky and Sapatinas [19] to bound T' and T''. Let j_A be the integer such that $2^{j_A} > \left(\frac{n}{\log n}\right)^{\frac{1}{2s+1}} > 2^{j_A - 1}$ (note that given our assumptions $j_0 \le j_A \le j_1$ for all sufficiently large n), then T' can be partitioned as $T' = T_1' + T_2'$, where the first component is calculated over the set of indices $j_0 \le j \le j_A$ and the second component over $j_A + 1 \le j \le j_1$. Hence, using Lemma 4.1 we obtain

$$T_1' \le C \sum_{j=j_0}^{j_A} \frac{2^j}{n} = O\left(2^{j_A} n^{-1}\right) = O\left(\left(\frac{n}{\log n} \right)^{\frac{1}{2s+1}} n^{-1} \right) \le O\left(n^{-\frac{2s}{2s+1}} \right). \tag{4.7}$$

To obtain a bound for T_2', we will use that if $f \in F_{p,q}^s(A)$, then for some constant C, dependent on s, p, q and $A > 0$ only, we have that

$$\sum_{k=0}^{2^j-1} b_{j,k}^2 \leq C 2^{-2js^*}, \tag{4.8}$$

for $1 \leq p \leq 2$, where $s^* = s + \frac{1}{2} - \frac{1}{p}$. Taking into account that $I\left(|b_{j,k}| > \frac{\xi_{j,n}}{2}\right) \leq \frac{4}{\xi_{j,n}^2}|b_{j,k}|^2$, we get

$$\begin{aligned} T_2' &\leq \frac{C}{n}\sum_{j=j_A}^{j_1}\sum_{k=0}^{2^j-1}\frac{4}{\xi_{j,n}^2}|b_{j,k}|^2 \leq \frac{C\left(\|f\|_\infty\right)2^{-2s^*j_A}}{\left(\sqrt{\delta\log n}+\|\psi\|_\infty \delta n^{\frac{-s}{2s+1}}(\log n)^{\frac{4s+1}{4s+2}}\right)^2}\sum_{j=j_A}^{j_1}\sum_{k=0}^{2^j-1}2^{2js^*}|b_{j,k}|^2 \\ &\leq O\left(2^{-2s^*j_A}\right) = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s^*}{2s+1}}\right), \end{aligned}$$

where we used the fact that $\sqrt{\delta\log n}+\|\psi\|_\infty \delta n^{\frac{-s}{2s+1}}(\log n)^{\frac{4s+1}{4s+2}} \to +\infty$ when $n \to +\infty$. Now remark that if $p = 2$ then $s^* = s$ and thus

$$T_2' = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s^*}{2s+1}}\right) = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right). \tag{4.9}$$

For the case $1 \leq p < 2$, the repeated use of the fact that if $B, D > 0$ then $I(|b_{j,k}| > B + D) \leq I(|b_{j,k}| > B)$, enables us to obtain that

$$\begin{aligned} T_2' &\leq \frac{C}{n}\sum_{j=j_A}^{j_1}\sum_{k=0}^{2^j-1} I\left(|b_{j,k}| > \frac{\xi_{j,n}}{2}\right) \leq \frac{C}{n}\sum_{j=j_A}^{j_1}\sum_{k=0}^{2^j-1}|b_{j,k}|^{-p}|b_{j,k}|^p I\left(|b_{j,k}|^{-p} < \left(2\|f\|_\infty\sqrt{\delta}\sqrt{\frac{\log n}{n}}\right)^{-p}\right) \\ &\leq C\sum_{j=j_A}^{j_1}\sum_{k=0}^{2^j-1}\frac{1}{n}\left(2\|f\|_\infty\sqrt{\delta}\sqrt{\frac{\log n}{n}}\right)^{-p}|b_{j,k}|^p. \end{aligned}$$

Since $f \in F_{p,q}^s(A)$ it follows that there exists a constant C depending only on p, q, s and A such that

$$\sum_{k=0}^{2^j-1}|b_{j,k}|^p \leq C2^{-pjs^*}, \tag{4.10}$$

where $s^* = s + \frac{1}{2} - \frac{1}{p}$ as before. By (4.10) we get

$$\begin{aligned} T_2' &\leq (\log n)\, C\left(\|f\|_\infty, \delta, p\right)\sum_{j=j_A}^{j_1}\sum_{k=0}^{2^j-1}\frac{(\log n)^{-\frac{p}{2}}}{n^{1-\frac{p}{2}}}|b_{j,k}|^p \leq C\left(\|f\|_\infty, \delta, p\right)\frac{(\log n)^{1-\frac{p}{2}}}{n^{1-\frac{p}{2}}}\sum_{j=j_A}^{j_1}C2^{-pjs^*} \\ &= O\left(\frac{(\log n)^{1-\frac{p}{2}}}{n^{1-\frac{p}{2}}}2^{-pj_As^*}\right) = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right). \end{aligned} \tag{4.11}$$

Hence, by (4.7), (4.9) and (4.11), $T' = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right)$.

Now, set j_A as before, then T'' can be split into $T'' = T_1'' + T_2''$, where the first component is calculated over the set of indices $j_0 \le j \le j_A$ and the second component over $j_A + 1 \le j \le j_1$. Then

$$T_1'' \le \sum_{j=j_0}^{j_A} \sum_{k=0}^{2^j-1} b_{j,k}^2 I\left(|b_{j,k}|^2 \le 32\left[4\|f\|_\infty^2\left(\sqrt{\frac{\delta\log n}{n}} + 2^{\frac{j}{2}}\|\psi\|_\infty \frac{\delta\log n}{n}\right)^2 + \frac{C_*^2}{n}\right]\right).$$

Using repeatedly that $(B+D)^2 \le 2\left(B^2+D^2\right)$ for $B, D \in \mathbb{R}$, we obtain the desired bound for T_1'':

$$\begin{aligned}
T_1'' &\le C\left(\|f\|_\infty\right)\sum_{j=j_0}^{j_A}\sum_{k=0}^{2^j-1}\left(\frac{\delta\log n}{n} + 2^j\|\psi\|_\infty^2\frac{\delta^2(\log n)^2}{n^2}\right) + C\left(C_*\right)\sum_{j=j_0}^{j_A}\sum_{k=0}^{2^j-1}\frac{1}{n}\\
&\le C\left(\|f\|_\infty,\delta,C_*\right)\frac{\log n}{n}2^{j_A} + C\left(\|f\|_\infty,\delta,\|\psi\|_\infty\right)\frac{(\log n)^2}{n^2}2^{2j_A}\\
&= O\left((\log n)^{\frac{2s}{2s+1}} n^{-\frac{2s}{2s+1}} + (\log n)^{\frac{4s}{2s+1}} n^{-\frac{4s}{2s+1}}\right) \le O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right). \qquad (4.12)
\end{aligned}$$

To bound T_2'', note that $T_2'' \le \sum_{j=j_A+1}^{j_1}\sum_{k=0}^{2^j-1} b_{j,k}^2 = O\left(2^{-2j_A s^*}\right) = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s^*}{2s+1}}\right)$, where we have used the condition (4.8). Now remark that if $p=2$ then $s^* = s$ and thus

$$T_2'' = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s^*}{2s+1}}\right) = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right). \qquad (4.13)$$

If $1 \le p < 2$,

$$\begin{aligned}
T_2'' &= \sum_{j=j_A+1}^{j_1}\sum_{k=0}^{2^j-1}|b_{j,k}|^{2-p}|b_{j,k}|^p I\left(|b_{j,k}| \le 2\xi_{j,n}\right)\\
&\le \sum_{j=j_A+1}^{j_1}\sum_{k=0}^{2^j-1}\left(8\|f\|_\infty\sqrt{\frac{\delta\log n}{n}} + 8\|f\|_\infty 2^{\frac{j_1}{2}}\|\psi\|_\infty\frac{\delta\log n}{n} + 4\sqrt{\frac{\log n}{n}}\right)^{2-p}|b_{j,k}|^p\\
&\le \left(C\left(\|f\|_\infty,\|\psi\|_\infty,\delta\right)\right)^{2-p}\left(\sqrt{\frac{\log n}{n}}\right)^{2-p}\sum_{j=j_A+1}^{j_1}\sum_{k=0}^{2^j-1}|b_{j,k}|^p = O\left(\left(\frac{\log n}{n}\right)^{\frac{2-p}{2}} 2^{-pj_A s^*}\right)\\
&= O\left(\left(\frac{n}{\log n}\right)^{\frac{p}{2}-1-\frac{p\left(s+\frac{1}{2}-\frac{1}{p}\right)}{2s+1}}\right) = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right), \qquad (4.14)
\end{aligned}$$

where we have used condition (4.10) and the fact that $C_* \le \sqrt{\log n}$ for n sufficiently large, taking into account that the constant $C_* := C\left(C_1, C_2\right)$ does not depend on n. Hence, by (4.12), (4.13) and (4.14), $T'' = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right)$. Combining all terms in (4.5), we conclude that:

$$\mathbb{E}\left\|\beta - \widehat{\beta}_{\xi_{j,n}}\right\|_2^2 = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right).$$

This completes the proof. ■

Lemma 4.4 *Assume that $f \in F^s_{p,q}(M)$ with $s > \frac{1}{2} + \frac{1}{p}$ and $1 \le p \le 2$. Suppose that Assumptions 1, 2 and 3 hold. For any $n > 1$, define $j_0 = j_0(n)$ to be the integer such that $2^{j_0} > \log n \ge 2^{j_0 - 1}$, and $j_1 = j_1(n)$ to be the integer such that $2^{j_1} \ge \frac{n}{\log n} \ge 2^{j_1 - 1}$. Define the threshold $\widehat{\xi}_{j,n}$ as in (2.7) for some constants $\delta = 6$ and $b \in \left[\frac{3}{4}, 1\right)$. Let $\beta_{j,k} := \langle f, \psi_{j,k} \rangle$ and $\widehat{\beta}_{\widehat{\xi}_{j,n},(j,k)} := \delta_{\widehat{\xi}_{j,n}}\left(\widehat{\beta}_{j,k}\right)$ with $(j,k) \in \Lambda_{j_1}$ as in (2.5). Take $\beta = (\beta_{j,k})_{(j,k)\in\Lambda_{j_1}}$ and $\widehat{\beta}_{\widehat{\xi}_{j,n}} = \left(\widehat{\beta}_{\widehat{\xi}_{j,n},(j,k)}\right)_{(j,k)\in\Lambda_{j_1}}$. Then, if $\|f - f_n\|_\infty \le \frac{1}{4}\|f\|_\infty$ and $N_n \le \frac{\kappa}{(r+1)^2}\frac{n}{\log n}$, where κ is a numerical constant and r is the degree of the polynomials, there exists a constant $M_4 > 0$ such that for all sufficiently large n:*

$$\mathbb{E}\left\|\beta - \widehat{\beta}_{\widehat{\xi}_{j,n}}\right\|_2^2 := \mathbb{E}\left(\sum_{(j,k)\in\Lambda_{j_1}} \left|\delta_{\widehat{\xi}_{j,n}}\left(\widehat{\beta}_{j,k}\right) - \beta_{j,k}\right|^2\right) \le M_4\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}$$

uniformly over $F^s_{p,q}(M)$.

Proof. Recall that f_n is the L_2 orthogonal projection of f on the space $\mathcal{P}_n$ of piecewise polynomials of degree r on a dyadic partition with step 2^{-J_n}. The dimension of $\mathcal{P}_n$ is $N_n = (r+1)2^{J_n}$. Let $\widehat{f}_n$ similarly be the orthogonal projection of I_n on $\mathcal{P}_n$. By doing analogous work as the one done to obtain (4.5), we get that

$$\mathbb{E}\left\|\beta - \widehat{\beta}_{\widehat{\xi}_{j,n}}\right\|_2^2 := T_1 + T_2 + T_3, \tag{4.15}$$

where $T_1 = \sum_{k=0}^{2^{j_0}-1} \mathbb{E}\left(a_{j_0,k} - \widehat{a}_{j_0,k}\right)^2$ do not depend on $\widehat{\xi}_{j,n}$. Therefore, by (4.6), $T_1 = O\left(n^{-\frac{2s}{2s+1}}\right)$. For T_2 and T_3 we have that

$$T_2 = \sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} \mathbb{E}\left[\left(b_{j,k} - \widehat{b}_{j,k}\right)^2\left\{I\left(\left|\widehat{b}_{j,k}\right| > \widehat{\xi}_{j,n}\,, |b_{j,k}| > \frac{\widehat{\xi}_{j,n}}{2}\right) + I\left(\left|\widehat{b}_{j,k}\right| > \widehat{\xi}_{j,n}\,, |b_{j,k}| \le \frac{\widehat{\xi}_{j,n}}{2}\right)\right\}\right]$$

and

$$T_3 = \sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} b^2_{j,k}\left[P\left(\left|\widehat{b}_{j,k}\right| \le \widehat{\xi}_{j,n}\,, |b_{j,k}| \le 2\widehat{\xi}_{j,n}\right) + P\left(\left|\widehat{b}_{j,k}\right| \le \widehat{\xi}_{j,n}, |b_{j,k}| > 2\widehat{\xi}_{j,n}\right)\right].$$

Using the same decomposition as in the proof of Lemma (4.3) we get that

$$\begin{aligned} T_2 + T_3 \le &\sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} \mathbb{E}\left\{\left(b_{j,k} - \widehat{b}_{j,k}\right)^2 I\left(|b_{j,k}| > \frac{\widehat{\xi}_{j,n}}{2}\right)\right\} + \sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} \mathbb{E}\left\{b^2_{j,k} I\left(|b_{j,k}| \le 2\widehat{\xi}_{j,n}\right)\right\} \\ &+ 5\sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} \mathbb{E}\left\{\left(b_{j,k} - \widehat{b}_{j,k}\right)^2 I\left(\left|\widehat{b}_{j,k} - b_{j,k}\right| > \frac{\widehat{\xi}_{j,n}}{2}\right)\right\} := T' + T'' + T'''. \end{aligned}$$

Now we bound T'''. Using Cauchy-Schwarz inequality, one obtains

$$T''' \le 5\sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} \mathbb{E}^{\frac{1}{2}}\left[\left(b_{j,k} - \widehat{b}_{j,k}\right)^4\right] P^{\frac{1}{2}}\left(\left|\widehat{b}_{j,k} - b_{j,k}\right| > \frac{\widehat{\xi}_{j,n}}{2}\right),$$

From Lemma 4.2 we have that for any $y > 0$ the following exponential inequality holds:

$$P\left(\left|\widehat{b}_{j,k}-b_{j,k}\right| > 2\left\|f\right\|_{\infty}\left(\sqrt{\frac{y}{n}}+2^{\frac{j}{2}}\left\|\psi\right\|_{\infty}\frac{y}{n}\right)+\frac{C_*}{\sqrt{n}}\right) \leq 2e^{-y}. \tag{4.16}$$

As in Comte [7], let $\Theta_{n,b} = \left\{\left|\frac{\|\widehat{f}_n\|_{\infty}}{\|f\|_{\infty}}-1\right| < b\right\}$, with $b \in (0,1)$. Then, using that $P\left(\left|\widehat{b}_{j,k}-b_{j,k}\right| > B+D\right) \leq P\left(\left|\widehat{b}_{j,k}-b_{j,k}\right| > B\right)$ for $B, D > 0$, and taking $x_n = \frac{\delta \log n}{(1-b)^2}$, one gets

$$\begin{aligned}
&P\left(\left|\widehat{b}_{j,k}-b_{j,k}\right| > \frac{\widehat{\xi}_{j,n}}{2}\right) \leq P\left(\left|\widehat{b}_{j,k}-b_{j,k}\right| > 2\left\|\widehat{f}_n\right\|_{\infty}\left(\sqrt{\frac{x_n}{n}}+2^{\frac{j}{2}}\left\|\psi\right\|_{\infty}\frac{x_n}{n}\right)\right)\\
&\leq P\left(\left(\left|\widehat{b}_{j,k}-b_{j,k}\right| > 2\left\|\widehat{f}_n\right\|_{\infty}\left(\sqrt{\frac{x_n}{n}}+2^{\frac{j}{2}}\left\|\psi\right\|_{\infty}\frac{x_n}{n}\right)\right) \mid \Theta_{n,b}\right)P\left(\Theta_{n,b}\right)\\
&+P\left(\left(\left|\widehat{b}_{j,k}-b_{j,k}\right| > 2\left\|\widehat{f}_n\right\|_{\infty}\left(\sqrt{\frac{x_n}{n}}+2^{\frac{j}{2}}\left\|\psi\right\|_{\infty}\frac{x_n}{n}\right)\right) \mid \Theta^c_{n,b}\right)P\left(\Theta^c_{n,b}\right)\\
&:= P_1 P\left(\Theta_{n,b}\right)+P_2 P\left(\Theta^c_{n,b}\right).
\end{aligned}$$

In Comte [7] is proved that if $\|f-f_n\|_{\infty} \leq \frac{1}{4}\|f\|_{\infty}$ then $P\left(\Theta^c_{n,b}\right) \leq O\left(n^{-4}\right)$ for the choices of $1 \geq b \approx \frac{4}{6}\sqrt{\frac{5}{\pi}} = 0.841 \geq \frac{3}{4}$ and $N_n \leq \frac{1}{36(r+1)^2}\frac{n}{\log n}$, where $\kappa = \frac{1}{36}$ is the numerical constant in the hypothesis of our theorem. Following its proof it can be shown that this bound can be improved taking $\kappa = \frac{1}{36\left(\frac{7}{5}\right)}$ and b as before (see the three last equations of page 290 in [7]). With this selection of κ we obtain that $P\left(\Theta^c_{n,b}\right) \leq O\left(n^{-6}\right)$. Using that $P\left(\Theta_{n,b}\right) = O(1)$ and $P_2 = O(1)$, it only remains to bound the conditional probability P_1. On $\Theta_{n,b}$ the following inequalities hold:

$$(a)\quad \left\|\widehat{f}_n\right\|_{\infty} > (1-b)\left\|f\right\|_{\infty} \quad \text{and} \quad (b)\quad \left\|\widehat{f}_n\right\|_{\infty} < (1+b)\left\|f\right\|_{\infty}. \tag{4.17}$$

Then, using (4.17a) we get

$$P_1 \leq P\left(\left|\widehat{b}_{j,k}-b_{j,k}\right| > 2\left\|f\right\|_{\infty}\left(\sqrt{\frac{\delta \log n}{n}}+2^{\frac{j}{2}}\left\|\psi\right\|_{\infty}\frac{\delta \log n}{n}\right)\right) \leq 2e^{-\delta \log n} = 2n^{-\delta},$$

where the last inequality is obtained using (4.16) for $y = \delta \log n > 0$. Hence, using that $\delta = 6$, we get $P\left(\left|\widehat{b}_{j,k}-b_{j,k}\right| > \frac{\widehat{\xi}_{j,n}}{2}\right) \leq O\left(n^{-6}\right)$. Therefore $T''' \leq C \sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} n^{-2} \leq O\left(n^{-2}2^{j_1}\right) \leq O\left(n^{-\frac{2s}{2s+1}}\right)$.

Now we bound T'. Let j_A be the integer such that $2^{j_A} > \left(\frac{n}{\log n}\right)^{\frac{1}{2s+1}} > 2^{j_A-1}$, then $T' = T'_1 + T'_2$, where the first component is computed over the set of indices $j_0 \leq j \leq j_A$ and the second component over $j_A + 1 \leq j \leq j_1$. Hence, using Lemma 4.1 we obtain

$$T'_1 \leq \sum_{j=j_0}^{j_A}\sum_{k=0}^{2^j-1}\mathbb{E}\left(b_{j,k}-\widehat{b}_{j,k}\right)^2 \leq O\left(\frac{2^{j_A}}{n}\right) = O\left(\left(\frac{n}{\log n}\right)^{\frac{1}{2s+1}}n^{-1}\right) \leq O\left(n^{-\frac{2s}{2s+1}}\right).$$

To bound T_2', note that $T_2' := T_{2,1}' + T_{2,2}'$, where

$$T_{2,1}' = \sum_{j=j_A}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left\{ \left(b_{j,k} - \widehat{b}_{j,k}\right)^2 I\left(|b_{j,k}| > \frac{\widehat{\xi}_{j,n}}{2}, \Theta_{n,b}\right)\right\} \text{ and}$$

$$T_{2,2}' = \sum_{j=j_A}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left\{ \left(b_{j,k} - \widehat{b}_{j,k}\right)^2 I\left(|b_{j,k}| > \frac{\widehat{\xi}_{j,n}}{2}, \Theta_{n,b}^c\right)\right\}.$$

Using that on $\Theta_{n,b}$ inequality (4.17a) holds and following the same procedures as in the proof of Theorem 2.6, we get the desired bound for $T_{2,1}'$.

$$\begin{aligned} T_{2,1}' &\leq \frac{C}{n} \sum_{j=j_A}^{j_1} \sum_{k=0}^{2^j-1} I\left(|b_{j,k}| > 2(1-b)\|f\|_\infty \left(\sqrt{\frac{x_n}{n}} + 2^{\frac{j}{2}} \|\psi\|_\infty \frac{x_n}{n}\right)\right) \\ &\leq \frac{C}{4}\frac{1}{n} \sum_{j=j_A}^{j_1} \sum_{k=0}^{2^j-1} \frac{|b_{j,k}|^2}{\|f\|_\infty^2 \left(\sqrt{\frac{\delta \log n}{n}} + 2^{\frac{j}{2}} \|\psi\|_\infty \frac{\delta \log n}{(1-b)n}\right)^2} \\ &\leq O\left(2^{-2s^* j_A}\right) = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s^*}{2s+1}}\right), \end{aligned} \tag{4.18}$$

where we have used that $\sqrt{\delta \log n} + \|\psi\|_\infty (1-b)^{-1} \delta n^{\frac{-s}{2s+1}} (\log n)^{\frac{4s+1}{4s+2}} \to +\infty$ when $n \to +\infty$ and that condition (4.8) is satisfied. Now remark that if $p = 2$ then $s^* = s$ and thus

$$T_{2,1}' = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right). \tag{4.19}$$

For the case $1 \leq p < 2$, from (4.18) we have that

$$\begin{aligned} T_{2,1}' &\leq \frac{C}{n} \sum_{j=j_A}^{j_1} \sum_{k=0}^{2^j-1} I\left(|b_{j,k}| > 2\|f\|_\infty \left(\sqrt{\frac{\delta \log n}{n}} + 2^{\frac{j}{2}} \|\psi\|_\infty \frac{\delta \log n}{(1-b)\,n}\right)\right) \\ &\leq \frac{C}{n} \sum_{j=j_A}^{j_1} \sum_{k=0}^{2^j-1} |b_{j,k}|^{-p} |b_{j,k}|^p I\left(|b_{j,k}|^{-p} < \left(2\|f\|_\infty \sqrt{\delta \frac{\log n}{n}}\right)^{-p}\right) \\ &\leq (\log n)\, C(\|f\|_\infty, \delta, p) \frac{(\log n)^{-\frac{p}{2}}}{n^{1-\frac{p}{2}}} \sum_{j=j_A}^{j_1} \sum_{k=0}^{2^j-1} |b_{j,k}|^p = O\left(\frac{(\log n)^{1-\frac{p}{2}}}{n^{1-\frac{p}{2}}} 2^{-p j_A s^*}\right) \\ &\leq O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right), \end{aligned} \tag{4.20}$$

where we have used condition (4.10). Hence $T_{2,1}' = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right)$.

Now we bound $T'_{2,2}$. Using Cauchy-Schwarz inequality, we have

$$T'_{2,2} \leq C \sum_{j=j_A}^{j_1} \sum_{k=0}^{2^j-1} n P^{\frac{1}{2}} \left(|b_{j,k}| > 2 \left\| \widehat{f}_n \right\|_\infty \left(\sqrt{\frac{x_n}{n}} + 2^{\frac{j}{2}} \|\psi\|_\infty \frac{x_n}{n} \right) + \sqrt{\frac{\log n}{n}} \mid \Theta^c_{n,b} \right) P^{\frac{1}{2}} \left(\Theta^c_{n,b}\right)$$

$$\leq C \sum_{j=j_A}^{j_1} \sum_{k=0}^{2^j-1} n P^{\frac{1}{2}} \left(\Theta^c_{n,b}\right) \leq C \sum_{j=j_A}^{j_1} \sum_{k=0}^{2^j-1} n^{-2} \leq O\left(\frac{2^{j_1}}{n^2}\right) \leq O\left(n^{-\frac{2s}{2s+1}}\right), \tag{4.21}$$

where we have used that $\mathbb{E}\left\{ \left(b_{j,k} - \widehat{b}_{j,k}\right)^4 \right\} = O\left(n^2\right)$ and that $P\left(\Theta^c_{n,b}\right) \leq O\left(n^{-6}\right)$. Then, putting together (4.19), (4.20) and (4.21), we obtain that $T'_2 = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right)$.

Now we bound T''. Set j_A as before, then $T'' = T''_1 + T''_2$, where the first component is calculated over the set of indices $j_0 \leq j \leq j_A$ and the second component over $j_A + 1 \leq j \leq j_1$. Recall that $x_n = \frac{\delta \log n}{(1-b)^2}$, then $T''_1 \leq T''_{1,1} + T''_{1,2}$, where

$$T''_{1,1} = \sum_{j=j_0}^{j_A} \sum_{k=0}^{2^j-1} b^2_{j,k} P \left(|b_{j,k}| \leq 4 \left[2(1+b) \|f\|_\infty \left(\sqrt{\frac{x_n}{n}} + 2^{\frac{j}{2}} \|\psi\|_\infty \frac{x_n}{n} \right) + \sqrt{\frac{\log n}{n}} \right] \right) \quad \text{and}$$

$$T''_{1,2} = \sum_{j=j_0}^{j_A} \sum_{k=0}^{2^j-1} b^2_{j,k} P\left(\Theta^c_{n,b}\right),$$

where we have used that given $\Theta_{n,b}$ inequality (4.17b) holds. For $T''_{1,1}$ we have

$$T''_{1,1} = \sum_{j=j_0}^{j_A} \sum_{k=0}^{2^j-1} b^2_{j,k} P \left(|b_{j,k}|^2 \leq 16 \left[2(1+b) \|f\|_\infty \left(\sqrt{\frac{x_n}{n}} + 2^{\frac{j}{2}} \|\psi\|_\infty \frac{x_n}{n} \right) + \sqrt{\frac{\log n}{n}} \right]^2 \right)$$

$$\leq 16 \sum_{j=j_0}^{j_A} \sum_{k=0}^{2^j-1} \left[2(1+b) \|f\|_\infty \left(\sqrt{\frac{\delta \log n}{(1-b)^2 n}} + 2^{\frac{j}{2}} \|\psi\|_\infty \frac{\delta \log n}{(1-b)^2 n} \right) + \sqrt{\frac{\log n}{n}} \right]^2$$

$$\leq C \sum_{j=j_0}^{j_A} \sum_{k=0}^{2^j-1} \left(C\left(\|f\|_\infty, b\right) \left(\frac{\delta \log n}{(1-b)^2 n} + \|\psi\|^2_\infty \frac{\delta^2 \log n}{(1-b)^4 n} \right) + \frac{\log n}{n} \right)$$

$$= O\left(\left(\frac{n}{\log n} \right)^{-\frac{2s}{2s+1}} \right), \tag{4.22}$$

where we have used repeatedly that $(B+D)^2 \leq 2\left(B^2 + D^2\right)$ for all $B, D \in \mathbb{R}$.

To bound $T''_{1,2}$ we use again that $P\left(\Theta^c_{n,b}\right) \leq O\left(n^{-6}\right)$ and that condition (4.8) is satisfied. Then

$$T''_{1,2} \leq \sum_{j=j_0}^{j_A} \sum_{k=0}^{2^j-1} b^2_{j,k} n^{-6} \leq n^{-6} \sum_{j=j_0}^{j_A} C 2^{-2js^*} = O\left(n^{-6} 2^{-2j_0 s^*}\right) \leq O\left(n^{-1}\right). \tag{4.23}$$

Hence, by (4.22) and (4.23), $T_1'' = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right)$. Now we bound T_2''.

$$T_2'' \leq \sum_{j=j_A+1}^{j_1} \sum_{k=0}^{2^j-1} b_{j,k}^2 P\left(|b_{j,k}| \leq 2\widehat{\xi}_{j,n}\right) \leq \sum_{j=j_A+1}^{j_1} \sum_{k=0}^{2^j-1} b_{j,k}^2 = O\left(2^{-2j_A s^*}\right) = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s^*}{2s+1}}\right),$$

where we have used again the condition (4.8). Now remark that if $p = 2$ then $s^* = s$ and thus $T_2'' = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s^*}{2s+1}}\right) = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right)$. For $1 \leq p < 2$, we proceed as follows.

$$\begin{aligned}
T_2'' &= \sum_{j=j_A+1}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left[b_{j,k}^2 I\left(|b_{j,k}| \leq 2\widehat{\xi}_{j,n}\,, \Theta_{n,b}\right) + b_{j,k}^2 I\left(|b_{j,k}| \leq 2\widehat{\xi}_{j,n}\,, \Theta_{n,b}^c\right)\right] \\
&\leq \sum_{j=j_A+1}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left[b_{j,k}^2 I\left(|b_{j,k}| \leq 4\left(2(1+b)\|f\|_\infty\left(\sqrt{\frac{x_n}{n}} + 2^{\frac{j}{2}}\|\psi\|_\infty \frac{x_n}{n}\right) + \sqrt{\frac{\log n}{n}}\right)\right)\right] \\
&+ \sum_{j=j_A+1}^{j_1} \sum_{k=0}^{2^j-1} b_{j,k}^2 P\left(|b_{j,k}| \leq \left(8\left\|\widehat{f}_n\right\|_\infty\left(\sqrt{\frac{x_n}{n}} + 2^{\frac{j}{2}}\|\psi\|_\infty \frac{x_n}{n}\right) + 4\sqrt{\frac{\log n}{n}}\right) \mid \Theta_{n,b}^c\right) P\left(\Theta_{n,b}^c\right) \\
&:= T_{2,1}'' + T_{2,2}'',
\end{aligned}$$

where we have used that on $\Theta_{n,b}$ inequality (4.17b) holds. Now we bound $T_{2,1}''$.

$$\begin{aligned}
T_{2,1}'' &\leq \sum_{j=j_A+1}^{j_1} \sum_{k=0}^{2^j-1} |b_{j,k}|^{2-p} |b_{j,k}|^p\, I\left(|b_{j,k}| \leq 8(1+b)\|f\|_\infty\left(\sqrt{\frac{x_n}{n}} + 2^{\frac{j}{2}}\|\psi\|_\infty \frac{x_n}{n}\right) + 4\sqrt{\frac{\log n}{n}}\right) \\
&\leq \sum_{j=j_A+1}^{j_1} \sum_{k=0}^{2^j-1} \left(8(1+b)\|f\|_\infty\left(\sqrt{\frac{\delta \log n}{(1-b)^2 n}} + 2^{\frac{j_1}{2}} \frac{\|\psi\|_\infty \delta \log n}{(1-b)^2 n}\right) + 4\sqrt{\frac{\log n}{n}}\right)^{2-p} |b_{j,k}|^p \\
&\leq C\left(\|f\|_\infty, b, \delta, \|\psi\|_\infty\right)^{2-p}\left(\sqrt{\frac{\log n}{n}}\right)^{2-p} \sum_{j=j_A+1}^{j_1} \sum_{k=0}^{2^j-1} |b_{j,k}|^p \leq O\left(\left(\frac{\log n}{n}\right)^{\frac{2-p}{2}} 2^{-pj_A s^*}\right) \\
&= O\left(\left(\frac{\log n}{n}\right)^{\frac{2-p}{2}} \left(\frac{n}{\log n}\right)^{-\frac{p\left(s+\frac{1}{2}-\frac{1}{p}\right)}{2s+1}}\right) = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right). \qquad (4.24)
\end{aligned}$$

where we have used that condition (4.10) is satisfied. To bound $T_{2,2}''$ we use again that $P\left(\Theta_{n,b}^c\right) \leq O\left(n^{-6}\right)$ and that condition (4.8) also holds. Then, from (4.23) we get

$$T_{2,2}'' \leq \sum_{j=j_A+1}^{j_1} \sum_{k=0}^{2^j-1} b_{j,k}^2 P\left(\Theta_{n,b}^c\right) \leq n^{-6} \sum_{j=j_A+1}^{j_1} C 2^{-2js^*} = O\left(n^{-6} 2^{-2j_A s^*}\right) \leq O\left(n^{-1}\right). \qquad (4.25)$$

Hence, by (4.24) and (4.25), $T'' = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right)$. Combining all terms in (4.15), we conclude that:

$$\mathbb{E}\left\|\beta - \widehat{\beta}_{\widehat{\xi}_{j,n}}\right\|_2^2 = O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right).$$

This completes the proof. ■

4.2 Proof of Theorem 2.5

First, one needs the following proposition.

Proposition 4.5 *Let $\beta_{j,k} = \langle f, \psi_{j,k} \rangle$ and $\widehat{\beta}_{j,k} = \langle I_n, \psi_{j,k} \rangle$ with $(j,k) \in \Lambda_{j_1}$. Suppose that $f \in F^s_{p,q}(M)$ with $s > 1/p$ and $1 \leq p \leq 2$. Let $M_1 > 0$ be a constant such that $M_1^{-1} \leq f \leq M_1$ (see Lemma 2.1). Let $\epsilon_{j_1} = 2M_1^2 e^{2\gamma_{j_1}+1} D_{j_1} A_{j_1}$. If $\epsilon_{j_1} \leq 1$, then there exists $\theta^*_{j_1} \in \mathbb{R}^{\#\Lambda_{j_1}}$ such that:*

$$\left\langle f_{j_1,\theta^*_{j_1}}, \psi_{j,k} \right\rangle = \langle f, \psi_{j,k} \rangle = \beta_{j,k} \text{ for all } (j,k) \in \Lambda_{j_1}$$

Moreover, the following inequality holds (approximation error)

$$\Delta\left(f; f_{j_1,\theta^*_{j_1}}\right) \leq \frac{M_1}{2} e^{\gamma_{j_1}} D^2_{j_1}.$$

Suppose that Assumptions 1 and 2 hold. Let $\eta_{j_1,n} = 4M_1^2 e^{2\gamma_{j_1}+2\epsilon_{j_1}+2} A^2_{j_1} \frac{\#\Lambda_{j_1}}{n}$. Then, for every $\lambda > 0$ such that $\lambda \leq \eta^{-1}_{j_1,n}$ there exists a set $\Omega_{n,1}$ of probability less than $M_2\lambda^{-1}$, where M_2 is the constant defined in Lemma 4.1, such that outside the set $\Omega_{n,1}$ there exists some $\widehat{\theta}_n \in \mathbb{R}^{\#\Lambda_{j_1}}$ which satisfies:

$$\left\langle f_{j_1,\widehat{\theta}_n}, \psi_{j,k} \right\rangle = \langle I_n, \psi_{j,k} \rangle = \widehat{\beta}_{j,k} \text{ for all } (j,k) \in \Lambda_{j_1}.$$

Moreover, outside the set $\Omega_{n,1}$, the following inequality holds (estimation error)

$$\Delta\left(f_{j_1,\theta^*_{j_1}}; f_{j_1,\widehat{\theta}_n}\right) \leq 2M_1 e^{\gamma_{j_1}+\epsilon_{j_1}+1} M_2 \lambda \frac{\#\Lambda_{j_1}}{n}.$$

Proof. Approximation error: Recall that $\beta_{j,k} = \langle f, \psi_{j,k} \rangle$ and let $\beta = (\beta_{j,k})_{(j,k)\in\Lambda_{j_1}}$. Define by $g_{j_1} = \sum_{(j,k)\in\Lambda_{j_1}} \theta_{j,k}\psi_{j,k}$ an approximation of $g = \log(f)$ and let $\beta_{0,(j,k)} = \left\langle f_{j_1,\theta_{j_1}}, \psi_{j,k} \right\rangle = \langle \exp(g_{j_1}), \psi_{j,k} \rangle$ with $\theta_{j_1} = (\theta_{j,k})_{(j,k)\in\Lambda_{j_1}}$ and $\beta_0 = \left(\beta_{0,(j,k)}\right)_{(j,k)\in\Lambda_{j_1}}$. Observe that the coefficients $\beta_{j,k} - \beta_{0,(j,k)}$, $(j,k) \in \Lambda_{j_1}$, are the coefficients of the orthonormal projection of $f - f_{j_1,\theta_{j_1}}$ onto V_j. Hence by Bessel's inequality, $\|\beta - \beta_0\|_2^2 \leq \left\| f - f_{j_1,\theta_{j_1}} \right\|^2_{L_2}$. Using Lemma 2.1 and Lemma 2 in Barron and Sheu [2], we get that:

$$\begin{aligned}
\|\beta - \beta_0\|_2^2 &\leq \int \left(f - f_{j_1,\theta_{j_1}}\right)^2 d\mu \leq M_1 \int \frac{\left(f - f_{j_1,\theta_{j_1}}\right)^2}{f} d\mu \\
&\leq M_1 e^{2\left\|\log\left(\frac{f}{f_{j_1,\theta_{j_1}}}\right)\right\|_\infty} \int f \left(\log\left(\frac{f}{f_{j_1,\theta_{j_1}}}\right)\right)^2 d\mu \\
&\leq M_1^2 e^{2\|g - g_{j_1}\|_\infty} \|g - g_{j_1}\|^2_{L_2} = M_1^2 e^{2\gamma_{j_1}} D^2_{j_1}.
\end{aligned}$$

Then, one can easily check that $b = e^{\left(\left\|\log\left(f_{j_1,\theta_{j_1}}\right)\right\|_\infty\right)} \leq M_1 e^{\gamma_{j_1}}$. Thus the assumption that $\epsilon_{j_1} \leq 1$ implies that the inequality $\|\beta - \beta_0\|_2 \leq M_1 e^{\gamma_{j_1}} D_{j_1} \leq \frac{1}{2beA_{j_1}}$ is satisfied. Hence, Lemma 2.4 can be applied with $\theta_0 = \theta_{j_1}$, $\widetilde{\beta} = \beta$ and $b = \exp\left(\left\|\log\left(f_{j_1,\theta_{j_1}}\right)\right\|_\infty\right)$, which implies that there exists $\theta^*_{j_1} = \theta(\beta)$ such that $\left\langle f_{j_1,\theta^*_{j_1}}, \psi_{j,k}\right\rangle = \beta_{j,k}$ for all $(j,k) \in \Lambda_{j_1}$.

By the Pythagorian-like relationship (2.2), we obtain that $\Delta\left(f; f_{j_1,\theta^*_{j_1}}\right) \leq \Delta\left(f; f_{j_1,\theta_{j_1}}\right)$. Now we use a result wich states that if f and g are two functions in $L_2([0,1])$ such that $\log\left(\frac{f}{g}\right)$ is bounded. Then $\Delta(f;g) \leq \frac{1}{2}e^{\left\|\log\left(\frac{f}{g}\right)\right\|_\infty} \int_0^1 f\left(\log\left(\frac{f}{g}\right)\right)^2 d\mu$, where μ denotes the Lebesgue measure on $[0,1]$.(see Lemma A.1 in Antoniadis and Bigot [1]). Hence, it follows that

$$\begin{aligned}\Delta\left(f; f_{j_1,\theta^*_{j_1}}\right) &\leq \frac{1}{2}e^{\left\|\log\left(\frac{f}{f_{j_1,\theta_{j_1}}}\right)\right\|_\infty} \int f\left(\log\left(\frac{f}{f_{j_1,\theta_{j_1}}}\right)\right)^2 d\mu \\ &= \frac{M_1}{2}e^{\|g-g_{j_1}\|_\infty}\|g-g_{j_1}\|^2_{L_2} = \frac{M_1}{2}e^{\gamma_{j_1}}D^2_{j_1}.\end{aligned}$$

which completes the proof for the approximation error.

Estimation error: Applying again Lemma 2.4 with $\theta_0 = \theta^*_{j_1}$, $\beta_{0,(j,k)} = \langle f_{j_1,\theta_0}, \psi_{j,k}\rangle = \beta_{j,k}$, $\widetilde{\beta} = \widehat{\beta}$, where $\widehat{\beta} = \left(\widehat{\beta}_{j,k}\right)_{(j,k)\in\Lambda_{j_1}}$, and $b = \exp\left(\left\|\log\left(f_{j_1,\theta^*_{j_1}}\right)\right\|_\infty\right)$ we obtain that if $\left\|\widehat{\beta} - \beta\right\|_2 \leq \frac{1}{2ebA_{j_1}}$ with $\beta = (\beta_{j,k})_{(j,k)\in\Lambda_{j_1}}$ then there exists $\widehat{\theta}_n = \theta\left(\widehat{\beta}\right)$ such that $\left\langle f_{j_1,\widehat{\theta}_n}, \psi_{j,k}\right\rangle = \widehat{\beta}_{j,k}$ for all $(j,k) \in \Lambda_{j_1}$.

Hence, it remains to prove that our assumptions imply that the event $\left\|\widehat{\beta} - \beta\right\|_2 \leq \frac{1}{2ebA_{j_1}}$ holds with probability $1 - M_2\lambda^{-1}$. First remark that $b \leq M_1 e^{\gamma_{j_1}+\epsilon_{j_1}}$ and that by Markov's inequality and Lemma 4.1 we obtain that for any $\lambda > 0$, $P\left(\left\|\widehat{\beta} - \beta\right\|_2^2 \geq \lambda\frac{\#\Lambda_{j_1}}{n}\right) \leq \frac{1}{\lambda}\frac{n}{\#\Lambda_{j_1}}\mathbb{E}\left\|\widehat{\beta} - \beta\right\|_2^2 \leq M_2\lambda^{-1}$. Hence, outside a set $\Omega_{n,1}$ of probability less than $M_2\lambda^{-1}$ then $\left\|\widehat{\beta} - \beta\right\|_2^2 \leq \lambda\frac{\#\Lambda_{j_1}}{n}$. Therefore, the condition $\left\|\widehat{\beta} - \beta\right\|_2 \leq \frac{1}{2ebA_{j_1}}$ holds if $\left(\lambda\frac{\#\Lambda_{j_1}}{n}\right)^{\frac{1}{2}} \leq \frac{1}{2ebA_{j_1}}$, which is equivalent to $4e^2b^2A^2_{j_1}\lambda\frac{\#\Lambda_{j_1}}{n} \leq 1$. This last inequality is true if $\eta_{j_1,n} = 4M_1^2 e^{2\gamma_{j_1}+2\epsilon_{j_1}+2}A^2_{j_1}\frac{\#\Lambda_{j_1}}{n} \leq \frac{1}{\lambda}$, using that $b^2 \leq M_1^2 e^{2\gamma_{j_1}+2\epsilon_{j_1}}$.

Hence, outside the set $\Omega_{n,1}$, our assumptions imply that there exists $\widehat{\theta}_n = \theta\left(\widehat{\beta}\right)$ such that $\left\langle f_{j_1,\widehat{\theta}_n}, \psi_{j,k}\right\rangle = \widehat{\beta}_{j,k}$ for all $(j,k) \in \Lambda_{j_1}$. Finally, outside the set $\Omega_{n,1}$, by using the bound given in Lemma 2.4, one obtains the following inequality for the estimation error

$$\Delta\left(f_{j_1,\theta^*_{j_1}}; f_{j_1,\widehat{\theta}_n}\right) \leq 2M_1 e^{\gamma_{j_1}+\epsilon_{j_1}+1}\lambda\frac{\#\Lambda_{j_1}}{n}.$$

which completes the proof of Proposition 4.5. ■

Our assumptions on $j_1(n)$ imply that $\frac{1}{2}n^{\frac{1}{2s+1}} \leq 2^{j_1(n)} \leq n^{\frac{1}{2s+1}}$. Therefore, using Lemma 2.2, one has that for all $f \in F^s_{2,2}(M)$ with $s > 1/2$

$$\gamma_{j_1(n)} \leq Cn^{\frac{1-2s}{2(2s+1)}}, \quad A_{j_1(n)} \leq Cn^{\frac{1}{2(2s+1)}}, \quad D_{j_1(n)} \leq Cn^{-\frac{s}{2s+1}},$$

where C denotes constants not depending on $g = \log(f)$. Hence, $\lim_{n\to+\infty}\epsilon_{j_1(n)} = \lim_{n\to+\infty} 2M_1^2 e^{2\gamma_{j_1(n)}+1}A_{j_1(n)}D_{j_1(n)} =$ 0,uniformly over $F^s_{2,2}(M)$ for $s > 1/2$. For all sufficiently large n, $\epsilon_{j_1(n)} \leq 1$ and thus, using Proposition 4.5, there exists $\theta^*_{j_1(n)} \in \mathbb{R}^{\#\Lambda_{j_1(n)}}$ such that

$$\Delta\left(f; f_{j,\theta^*_{j_1(n)}}\right) \leq \frac{M_1}{2} e^{\gamma_{j_1(n)}} D^2_{j_1(n)} \leq C n^{-\frac{2s}{2s+1}} \text{ for all } f \in F^s_{2,2}(M). \tag{4.26}$$

By the same arguments it follows that $\lim_{n\to+\infty} \eta_{j_1(n),n} = \lim_{n\to+\infty} 4M_1^2 e^{2\gamma_{j_1(n)}+2\epsilon_{j_1(n)}+2} A^2_{j_1(n)} \frac{\#\Lambda_{j_1(n)}}{n} = 0$, uniformly over $F^s_{2,2}(M)$ for $s > 1/2$. Now let $\lambda > 0$. The above result shows that for sufficiently large n, $\lambda \leq \eta^{-1}_{j_1(n),n}$, and thus using Proposition 4.5 it follows that there exists a set $\Omega_{n,1}$ of probability less than $M_2\lambda^{-1}$ such that outside this set there exists $\widehat{\theta}_n \in \mathbb{R}^{\#\Lambda_{j_1(n)}}$ which satisfies:

$$\Delta\left(f_{j_1(n),\theta^*_{j_1(n)}}; f_{j_1(n),\widehat{\theta}_n}\right) \leq 2M_1 e^{\gamma_{j_1(n)}+\epsilon_{j_1(n)}+1} M_2 \lambda \frac{\#\Lambda_{j_1(n)}}{n} \leq C\lambda n^{-\frac{2s}{2s+1}}, \tag{4.27}$$

for all $f \in F^s_{2,2}(M)$. Then, by the Pythagorian-like identity (2.2) it follows that outside the set $\Omega_{n,1}$

$$\Delta\left(f; f_{j_1(n),\widehat{\theta}_n}\right) = \Delta\left(f; f_{j_1(n),\theta^*_{j_1(n)}}\right) + \Delta\left(f_{j_1(n),\theta^*_{j_1(n)}}; f_{j_1(n),\widehat{\theta}_n}\right),$$

and thus Theorem 2.5 follows from inequalities (4.26) and (4.27).

4.3 Proof of Theorem 2.6

First, one needs the following proposition.

Proposition 4.6 *Let $\beta_{j,k} := \langle f, \psi_{j,k}\rangle$ and $\widehat{\beta}_{\xi_{j,n},(j,k)} := \delta_{\xi_{j,n}}\left(\widehat{\beta}_{j,k}\right)$ with $(j,k) \in \Lambda_{j_1}$. Assume that $f \in F^s_{p,q}(A)$ with $s > 1/p$ and $1 \leq p \leq 2$. Let $M_1 > 0$ be a constant such that $M_1^{-1} \leq f \leq M_1$ (see Lemma 2.1). Let $\epsilon_{j_1} = 2M_1^2 e^{2\gamma_{j_1}+1} D_{j_1} A_{j_1}$. If $\epsilon_{j_1} \leq 1$, then there exists $\theta^*_{j_1} \in \mathbb{R}^{\#\Lambda_{j_1}}$ such that:*

$$\left\langle f_{j_1,\theta^*_{j_1}}, \psi_{j,k}\right\rangle = \langle f, \psi_{j,k}\rangle = \beta_{j,k} \text{ for all } (j,k) \in \Lambda_{j_1}$$

Moreover, the following inequality holds (approximation error)

$$\Delta\left(f; f_{j_1,\theta^*_{j_1}}\right) \leq \frac{M_1}{2} e^{\gamma_{j_1}} D^2_{j_1}.$$

Suppose that Assumptions 1 and 2 hold. Let $\eta_{j_1,n} = 4M_1^2 e^{2\gamma_{j_1}+2\epsilon_{j_1}+2} A^2_{j_1}\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}$. Then, for every $\lambda > 0$ such that $\lambda \leq \eta^{-1}_{j_1,n}$ there exists a set $\Omega_{n,2}$ of probability less than $M_3\lambda^{-1}$, where M_3 is the constant defined in Lemma 4.3, such that outside the set $\Omega_{n,2}$ there exists some $\widehat{\theta}_n \in \mathbb{R}^{\#\Lambda_{j_1}}$ which satisfies:

$$\left\langle f^{HT}_{j_1,\widehat{\theta}_n,\xi_{j,n}}, \psi_{j,k}\right\rangle = \delta_{\xi_{j,n}}\left(\widehat{\beta}_{j,k}\right) = \widehat{\beta}_{\xi_{j,n},(j,k)} \text{ for all } (j,k) \in \Lambda_{j_1}.$$

Moreover, outside the set $\Omega_{n,2}$, the following inequality holds (estimation error)

$$\Delta\left(f_{j_1,\theta^*_{j_1}}; f^{HT}_{j_1,\widehat{\theta}_n,\xi_{j,n}}\right) \leq 2M_1 e^{\gamma_{j_1}+\epsilon_{j_1}+1} \lambda \left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}.$$

Proof. Approximation error: The proof is the same that the one of Proposition 4.5.
Estimation error: Applying Lemma 2.4 with $\theta_0 = \theta^*_{j_1}$, $\beta_{0,(j,k)} = \langle f_{j_1,\theta_0}, \psi_{j,k}\rangle = \beta_{j,k}$, $\widetilde{\beta} = \widehat{\beta}_{\xi_{j,n}}$, where $\widehat{\beta}_{\xi_{j,n}} = \left(\widehat{\beta}_{\xi_{j,n},(j,k)}\right)_{(j,k)\in\Lambda_{j_1}}$, and $b = \exp\left(\left\|\log\left(f_{j_1,\theta^*_{j_1}}\right)\right\|_\infty\right)$ we obtain that if $\left\|\widehat{\beta}_{\xi_{j,n}} - \beta\right\|_2 \le \frac{1}{2ebA_{j_1}}$ with $\beta = (\beta_{j,k})_{(j,k)\in\Lambda_{j_1}}$ then there exists $\widehat{\theta}_n = \theta\left(\widehat{\beta}_{\xi_{j,n}}\right)$ such that $\left\langle f^{HT}_{j_1,\widehat{\theta}_{j_1},\xi_{j,n}}, \psi_{j,k}\right\rangle = \widehat{\beta}_{\xi_{j,n},(j,k)}$ for all $(j,k)\in\Lambda_{j_1}$.

Hence, it remains to prove that our assumptions imply that the event $\left\|\widehat{\beta}_{\xi_{j,n}} - \beta\right\|_2 \le \frac{1}{2ebA_{j_1}}$ holds with probability $1 - M_3\lambda^{-1}$. First remark that $b \le M_1 e^{\gamma_{j_1}+\epsilon_{j_1}}$ and that by Markov's inequality and Lemma 4.3 we obtain that for any $\lambda > 0$,

$$P\left(\left\|\widehat{\beta}_{\xi_{j,n}} - \beta\right\|_2^2 \ge \lambda\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right) \le \frac{1}{\lambda}\left(\frac{n}{\log n}\right)^{\frac{2s}{2s+1}} \mathbb{E}\left\|\widehat{\beta}_{\xi_{j,n}} - \beta\right\|_2^2$$

$$\le \frac{M_3}{\lambda}\left(\frac{n}{\log n}\right)^{\frac{2s}{2s+1}}\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}} \le M_3\lambda^{-1}.$$

Hence, outside a set $\Omega_{n,2}$ of probability less than $M_3\lambda^{-1}$, it holds that $\left\|\widehat{\beta}_{\xi_{j,n}} - \beta\right\|_2^2 \le \lambda\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}$. Therefore, the condition $\left\|\widehat{\beta}_{\xi_{j,n}} - \beta\right\|_2 \le \frac{1}{2ebA_{j_1}}$ holds if $\left(\lambda\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right)^{\frac{1}{2}} \le \frac{1}{2ebA_{j_1}}$, which is equivalent to $4e^2b^2A^2_{j_1}\lambda\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}} \le 1$. Using that $b^2 \le M_1^2 e^{2\gamma_{j_1}+2\epsilon_{j_1}}$ the last inequality is true if $\eta_{j_1,n} = 4M_1^2 e^{2\gamma_{j_1}+2\epsilon_{j_1}+2}A^2_{j_1}\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}} \le \frac{1}{\lambda}$.

Hence, outside the set $\Omega_{n,2}$, our assumptions imply that there exists $\widehat{\theta}_n = \theta\left(\widehat{\beta}_{\xi_{j,n}}\right)$ such that $\left\langle f^{HT}_{j_1,\widehat{\theta}_n,\xi_{j,n}}, \psi_{j,k}\right\rangle = \widehat{\beta}_{\xi_{j,n},(j,k)}$ for all $(j,k)\in\Lambda_{j_1}$. Finally, outside the set $\Omega_{n,2}$, by using the bound given in Lemma 2.4, one obtains the following inequality for the estimation error

$$\Delta\left(f_{j_1,\theta^*_{j_1}}; f^{HT}_{j_1,\widehat{\theta}_n,\xi_{j,n}}\right) \le 2M_1 e^{\gamma_{j_1}+\epsilon_{j_1}+1}\lambda\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}},$$

which completes the proof of Proposition 4.6. ■

Our assumptions on $j_1(n)$ imply that $\frac{1}{2}\frac{n}{\log n} \le 2^{j_1(n)} \le \frac{n}{\log n}$. Therefore, using Lemma 2.2, one has that for all $f \in F^s_{p,q}(M)$ with $s > 1/p$,

$$\gamma_{j_1(n)} \le C\left(\frac{n}{\log n}\right)^{-\left(s-\frac{1}{p}\right)}, \quad A_{j_1(n)} \le \left(\frac{n}{\log n}\right)^{\frac{1}{2}}, \quad D_{j_1(n)} \le C\left(\frac{n}{\log n}\right)^{-s^*},$$

where C denotes constants not depending on $g = \log(f)$. Hence,

$$\lim_{n\to+\infty}\epsilon_{j_1(n)} = \lim_{n\to+\infty} 2M_1^2 e^{2\gamma_{j_1(n)}+1}A_{j_1(n)}D_{j_1(n)} = 0,$$

uniformly over $F^s_{p,q}(M)$ for $s > 1/p$. For all sufficiently large n, $\epsilon_{j_1(n)} \le 1$ and thus, using Proposition 4.6, there exists $\theta^*_{j_1(n)} \in \mathbb{R}^{\#\Lambda_{j_1(n)}}$ such that

$$\Delta\left(f; f_{j_1(n),\theta^*_{j_1(n)}}\right) \le \frac{M_1}{2}e^{\gamma_{j_1(n)}}D^2_{j_1(n)} \le C\left(\frac{n}{\log n}\right)^{-2s^*} \quad \text{for all } f \in F^s_{p,q}(M).$$

Now remark that if $p = 2$ then $s^* = s > 1$ (by assumption), thus

$$\Delta\left(f; f_{j_1(n),\theta^*_{j_1(n)}}\right) = O\left(\left(\frac{n}{\log n}\right)^{-2s}\right) \leq O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right).$$

If $1 \leq p < 2$ then one can check that condition $s > \frac{1}{2} + \frac{1}{p}$ implies that $2s^* > \frac{2s}{2s+1}$, hence

$$\Delta\left(f; f_{j_1(n),\theta^*_{j_1(n)}}\right) \leq O\left(\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}}\right). \tag{4.28}$$

By the same arguments it holds that

$$\lim_{n\to+\infty} \eta_{j_1(n),n} = \lim_{n\to+\infty} 4M_1^2 e^{2\left(\gamma_{j_1(n)}+\epsilon_{j_1(n)}+1\right)} A^2_{j_1(n)} \left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}} = 0,$$

uniformly over $F^s_{p,q}(M)$ for $s > 1/p$. Now let $\lambda > 0$. The above result shows that for sufficiently large n, $\lambda \leq \eta^{-1}_{j_1(n),n}$, and thus using Proposition 4.6 it follows that there exists a set $\Omega_{n,2}$ of probability less than $M_3\lambda^{-1}$ such that outside this set there exists $\widehat{\theta}_n \in \mathbb{R}^{\#\Lambda_{j_1(n)}}$ which satisfies:

$$\Delta\left(f_{j_1(n),\theta^*_{j_1(n)}}; f^{HT}_{j_1(n),\widehat{\theta}_n,\xi_{j,n}}\right) \leq 2M_1 e^{\gamma_{j_1(n)}+\epsilon_{j_1(n)}+1} \lambda \left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+1}} \tag{4.29}$$

for all $f \in F^s_{p,q}(M)$. Then, by the Pythagorian-like identity (2.2) it follows that outside the set $\Omega_{n,2}$,

$$\Delta\left(f; f^{HT}_{j_1(n),\widehat{\theta}_n,\xi_{j,n}}\right) = \Delta\left(f; f_{j_1(n),\theta^*_{j_1(n)}}\right) + \Delta\left(f_{j_1(n),\theta^*_{j_1(n)}}; f^{HT}_{j_1(n),\widehat{\theta}_n,\xi_{j,n}}\right),$$

and thus Theorem 2.6 follows from inequalities (4.28) and (4.29).

4.4 Proof of Theorem 2.7

The proof is analogous to the one of Theorem 2.6. It follows from Lemma 4.4.

References

[1] Antoniadis, A and Bigot, J. (2006). Poisson inverse problems. *Ann. Statist.,* **34**, 2132-2158.

[2] Barron, A. R. and Sheu, C. H. (1991). Approximation of density functions by sequences of exponential families. *Ann. Statist.,* **19**, 1347-1369.

[3] Bochner, S. (1932). Vorlesungen uber Fouriersche Integrale. Akademische Verlag-sgesellschaft, Leipzig.

[4] Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory.* New York: McGraw-Hill Inc.

[5] Bigot, J. and Van Bellegem, S. (2009) Log-density deconvolution by wavelet thresholding, *Scandinavian Journal of Statistics*, to be published.

[6] Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves: exponential bounds and rates of convergence, *Bernouilli*, **4** (3) 251-265. 329-475.

[7] Comte, F. (2001). Adaptive estimation of the spectrum of a stationary Gaussian sequence. *Bernoulli*, **7** (2), 267-298.

[8] Csiszár, I. (1975). I-divergence geometry of probability distributions and minimization problems. *Ann. Probab.,* **3**, 146–158.

[9] Daubechies, I. (1992). *Ten Lectures on Wavelets.* Philadelphia, PA: Society for Industrial and Applied Mathematics.

[10] Davies, R.B. (1973). Asymptotic inference in stationary Gaussian time-series. *Adv. Appl. Probab.*, **5**, 469-497.

[11] DeVore, R.A. and Lorentz, G.G. (1993). *Constructive Approximation.* Berlin: Springer-Verlag.

[12] Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425–55.

[13] Fryzlewicz, P., Nason, G.P., von Sachs, R. (2008). A Wavelet-Fisz Approach to Spectrum Estimation. *Journal of time series analysis.* Vol. **29**, No. 5, 868-880.

[14] Hardle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications.* Lecture Notes in Statistics 129. Springer.

[15] Koo, J.Y. (1999). Logspline Deconvolution in Besov Space. *Scandinavian Journal of Statistics*, **26**, 73–86.

[16] Loubes, J.-M. and Yan, Y. (2009). Penalized maximum likelihood estimation with l_1 penalty. *International Journal of Applied Mathematics and Statistics.* Vol **14**, No. J09, 35-46.

[17] Mallat, S. (1999). *A Wavelet Tour of Signal Processing.* 2nd ed. Academic Press, San Diego.

[18] Neumann, M. H. (1996). Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series. *Journal of time series analysis.* Vol. **17**, No. 6.

[19] Pensky, M. and Sapatinas, T. (2009). Functional deconvolution in a periodic setting: uniform case. *Ann. Statist.,* **37**, 73-104.

[20] Priestley, M. B. (1981). *Spectral Analysis and Time Series.* London: Academic Press.